REPORT RESUMES

ED 003 822 24

SEQUENTIAL CHARACTERISTICS OF LANGUAGE.

BY- JOHNSON, NEAL F.
OHIO STATE UNIV., COLUMBUS, RESEARCH FOUNDATION

REPORT NUMBER CRP-1996 PUB DATE 65
EDRS PRICE MF-$0.27 HC-$5.36 134P.

DESCRIPTORS- *SEQUENTIAL LEARNING, *GRAMMAR, *REINFORCEMENT, LANGUAGE INSTRUCTION, *LANGUAGE ABILITY, *LANGUAGE RESEARCH, STRUCTURAL ANALYSIS, COLUMBUS, OHIO

THIS RESEARCH EXAMINED THE WAYS SUBJECTS USE SUCH A STRUCTURAL DEVICE AS GRAMMAR IN THEIR ATTEMPTS TO RECALL OR PRODUCE BEHAVIOR SEQUENCES IN LANGUAGE INSTRUCTION. RESEARCH TECHNIQUES TO STUDY THE BEHAVIORAL CORRELATES OF THE UTILIZATION OF RULES IN THE GENERATION OF LANGUAGE BEHAVIOR WERE DEVELOPED. THESE TECHNIQUES WERE THEN USED TO STUDY THE WAY SUBJECTS USE THE RULES IN RESPONSE GENERATION. THE EXPERIMENTAL TASK WAS A PAIRED-ASSOCIATE LEARNING SITUATION IN WHICH SUBJECTS LEARNED SENTENCES AS RESPONSES TO DIGIT STIMULI. IT WAS CONCLUDED THAT THE AVAILABLE EVIDENCE INDICATES THAT SUBJECTS DO USE PHRASE-STRUCTURE RULES AS OPPOSED TO ITEM-TO-ITEM ASSOCIATIONS WHEN THEY ARE ATTEMPTING TO RECALL LANGUAGE SEQUENCES. (GD)

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

Sequential Characteristics of Language

Cooperative Research Project No. 1996

Neal F. Johnson

The Ohio State University

1965

The research reported herein was supported by the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education, and Welfare.

ED 003822

TABLE OF CONTENTS

CHAPTER		PAGE
I	INTRODUCTION	1
	Problems and Goals	1
	Related Research	1
II	METHODS AND PROCEDURES	20
	Experimental Task	21
	Subjects	22
	Scoring Techniques	23
III	PILOT EXPERIMENTS	25
	Experiment I	26
	Experiment II	30
	Experiment III	34
	Experiment IV	39
	Discussion of Pilot Experiments	43
IV	EXPERIMENTS TESTING THE SNOWBALL AND ACCUMULATING CONTEXT HYPOTHESES	48
	Experiment V	50
	Experiment VI	53
	Experiment VII	59
	Experiment VIII	61
	General Discussion	65
V	EXPERIMENTS TESTING VIABILITY OF THE DECODING-OPERATION HYPOTHESIS	69
	Experiment IX	71
	Experiment X	77
	Experiment XI	87
	Experiment XII	94
	Experiment XIII	102
	Experiment XIV	106
VI	CONCLUSIONS AND IMPLICATIONS	114
	References	122

CHAPTER I

Introduction

Problem and Goals

The problem area upon which the following research was focused concerned the serial ordering of language behavior. Specifically, the concern was with the way subjects (S̲s) used grammatical rules to produce language sequences. At the beginning of the project there seemed to be two problems. The first involved developing research techniques for the study of the behavioral correlates of grammatical rules, and the second involved determining the way S̲s use the rules in response generation.

The goals of the research were to study some of the variables which influence the learning of grammatical rules and to determine the extent to which these rules define functional units of language behavior. The project was based upon the general hypothesis that one role grammar plays in language behavior is that of providing "chunking" rules which defined functional decision units for a speaker.

Related Research

One of the first studies exploring the role of a structure-imposed contextual constraint on learning response sequences was an experiment by Miller and Selfridge (1950). In their experiment they manipulated contextual constraint by varying the degree to which their materials approximated the statistical structure of English. The sequences they used varied in length from 10 to 50 words and in approximation to English from zero-order approximations to seventh-order approximations, plus some textual material taken from books. The subjects (Ss) were presented with a list of sequences all of a particular length and approximation to English and then asked to recall all the words they could remember, irrespective of the order in which they occurred. Their results indicate that percentage of words recalled increased with statistical approximation to English and decreased with the length of the passage. In addition, they suggested the results indicate that the facilitating influence of the constraints on learning increased as the amount to be retained increased.

One of the interesting findings of the Miller and Selfridge experiment was that, even for the longest sequence, the degree of facilitation from the increasing degrees of contextual constraint seemed to reach an asymptote somewhere about a fifth-order approximation, with little improvement occurring beyond that point. The fact that there was a marked improvement in performance from a zero-order approximation to a second-order approximation, where the material still seemed meaningless, plus the fact that there was little improvement of passages from texts over fifth or seventh-order approximations, led Miller and Selfridge to suggest that their findings indicate the important distinction may not have been between meaningful vs. nonsense materials, but rather the degree to which the constraint allowed Ss to utilize the short-range associations they had established in their prior usage of English.

Marks and Jack (1952) performed an experiment which was a partial replication of Miller and Selfridge (1950) except they used a more rigorous scoring criterion. Miller and Selfridge scored _S_s for number of words recalled, irregardless of the order in which they were recalled. Marks and Jack, on the other hand, required the _S_s to recall all of the words from a sequence in their correct order. In addition, Miller and Selfridge used some partial sentences in their textual material, and Marks and Jack used only complete sentences. Where Miller and Selfridge showed almost no difference between fifth-order approximations to English and text, the data of Marks and Jack show a rather sizeable difference. On the basis of this they conclude, contrary to Miller and Selfridge, that "recall seems to be very much a function of the _a priori_ relative meaningfulness of the materials." In addition, Marks and Miller (1964) have demonstrated that if the words from grammatical and meaningful sentences were scrambled the sequences were learned more rapidly than if the original sentences were grammatical but meaningless.

The fact that in both the Marks and Jack (1952) and Miller and Selfridge (1950) experiments there were sizeable differences in degree of recall between adjacent orders of approximation at the lower end (e. g., between zero and first order, between first order and second order, etc.) suggests that if meaningfulness is the variable it is not dichotomous. In addition, the increase in apparent meaningfulness of second and third-order approximations over first-order approximations is so slight that it seems improbable it could account for the rather sizeable changes in performance related to these increasing degrees of approximation. That is, while meaningfulness may play a role, as suggested by Marks and Jack (1952), and Marks and Miller (1964), it also seems quite possible that there is an influence of constraints from other than semantic sources (e. g., Marks and Miller, 1964). Whether the constraint is semantic or syntactic, the conclusion of Miller and Selfridge (1950) that the role of the constraint is to preserve the short-range associations, and thus allow positive transfer, seems quite tenable.

In a pair of experiments by Aborn and Rubenstein (1952, 1954), Ss were presented with sequences of 30 to 32 nonsense syllables, after which they were asked to recall them. The nonsense syllables were sorted into four classes on the basis of the first letter of the nonsense syllable and there were four syllables per class, (i. e., 16 syllables in all). Sequential constraints were introduced to reduce the information per syllable. The constraints were based on rules which specified which classes of syllables could follow which other classes. Six passages were constructed from these nonsense syllables varying in mean rate of information per syllable from one to four bits. Before the experiments began the Ss memorized the 16 nonsense syllables and the classes to which they belonged. In addition, they were allowed to study the organization pattern of the passage to be learned for 10 minutes prior to learning the passage. Therefore, before the Ss were presented a sequence to be learned they had memorized the items which would be appearing in the sequence, plus the organizational characteristics or the form of the constraint. The results of their experiment indicate quite clearly that as the mean rate of information decreased, the number of syllables retained increased.

The results of the Aborn and Rubenstein experiments probably can be interpreted as demonstrating a usable source of contextual constraint that is nonsemantic. That is, while their various passages increased in contextual constraint, there is no sense in which one could say the passages increased in meaningfulness; for all of them were simply sequences of nonsense syllables. The experiment, however, raises some questions regarding the nature of the associative mechanisms established during the prior training which allowed the constraints to influence recall as they did. Where Miller and Selfridge (1950) interpreted the influence of the constraint in terms of its preserving short-range associations, the Rubenstein and Aborn experiments did not establish any specific short-range associations between items during the prior training. Therefore, while these experiments seem to support the conclusion of Miller and Selfridge that there is a nonsemantic source of constraint, it casts some doubt on the suggestion that it is simply associations between elements that are preserved.

An experiment by Miller (1958) raises another question regarding the interpretation that Miller and Selfridge (1950) give their data. He constructed a set of rules which would generate sequences consisting of four different letters, any one of which could be repeated any number of times. One group of Ss were presented with sequences, generated by the set of rules, which varied in length from four to seven letters. Another group of Ss learned strings of comparable length composed of the same letters but selected in a random order. The Ss were presented the nine strings they were to learn, each one appearing for five seconds. After they had seen all nine they were asked to reproduce as many as they could. They were given 10 such trials. The interesting difference between this experiment and the earlier ones is that the Ss had no experience with either the letters or the organization prior to the learning task. Nevertheless, the groups learning the structured sequences showed a consistent superiority over the group learning the random sequences, even on the first few trials. The fact that the group that was learning the structured sequences showed the consistent superiority, again raises the question as to whether the contextual constraint must come from previously established short-range associations. Certainly these did not exist in the Miller (1958) experiment and it is doubtful whether the item-to-item associations they imply were established in the Aborn and Rubenstein (1952, 1954) experiments.

While the above data offer support for the hypothesis that there is a source of variance attributable to a nonsemantic contextual constraint in response sequences, the nature of that constraint, and the way it should be characterized, is somewhat less apparent. On the surface, at least, it would seem the simplest account for this nonsemantic constraint would be in terms of increasing the predictibility of the items, and, consequently, either the usability of prior habits or a fewer number of new habits to be formed. Such a conception has been offered by Hockett (1955). He has suggested the possibility of representing the linguistic structure of a language in terms of a matrix in which are entered the transitional probabilities existing between words. The conception can be represented schematically as in Table 1.

The dimensions of the matrix would be all-of-the-words-in-English by all-of-the-words-in-English, and within the cells the numbers would indicate the probability that a word in a row margin would be followed by a word in a column margin. Such a model has a great deal of appeal from a psychological point of view because it presupposes only simple item-to-item associations and at the same time both accurately and completely represents the redundancy of the language. Assuming that the values within the cell were determined empirically, the probabilities should reflect the relative frequency of reinforcement and, consequently, could be translated as a measure of habit strength between the two items.

That conception would be all that is necessary to account for the results of the Miller and Selfridge (1950) experiment. The variation in statistical approximation to English among the sequences could be attained by constructing sequences which differed in mean transitional probability. Their data could be interpreted as reflecting those differing probabilities. The results

TABLE 1

	the	boy	red		turgid	blast	tooth
the	.00	.13	.10		.03	.04	.07
boy	.00	.00	.01		.02	.04	.01
red	.00	.00	.00		.03	.06	.00
turgid	.00	.02	.01		.00	.01	.00
blast	.05	.00	.03		.01	.00	.00
tooth	.00	.00	.00		.00	.00	.00

of the Miller (1958) experiment could be interpreted as suggesting it is easier to acquire or learn the habit matrix when some of the transitional probabilities are very high and the others are either zero or near zero, than is the case if they are all equal and above zero. A similar account can be given for the Aborn and Rubenstein (1952,1954) experiments. Their rules could be characterized by a matrix which had 16 nonsense syllables in the vertical and horizontal margins. The differences between their groups could be expressed in terms of the mean of the non-zero transitional probabilities.

It does seem the case, then, that the redundancy in at least some rule-governed language behavior can be expressed as a transitional probability matrix and the above data indicate that manipulating the values within the matrix results in predictable behavioral changes. If the only behavioral implications of "knowing-a-rule" are: 1) given knowledge of the _n_ preceding items, item _n_ + _1_ is more predictable than would be the case if either the rule were not known or the _n_ preceding items were not known; and 2) the only reason for formulating a characterization of the speaker's knowledge is to allow prediction of his future behavior, then a conception of the speaker's knowledge in terms of a habit or transitional probability matrix probably would be fairly adequate. If on the other hand, there are implications of rule-governed behavior beyond the simple frequency of its occurrence, then the matrix model will have to be elaborated.

There are several problems with attempting to characterize the structure of language in terms of a matrix of habits or transitional probabilities. The logical difficulties from the linguists' point of view have been reviewed by Chomsky (1957) and Lees (1964) and from a psychological point of view by Lashley (1951), Miller, Galanter, and Pribram (1960), and Osgood (1963). In addition, there are a number of empirical difficulties because there do seem to be some behavioral implications which are not reflected in the matrix of habits or transitional probabilities.

The first difficulty concerns the way the habits within the matrix are established. One of the interesting characteristics of learning a sample of structured material is that after the learning experience the Ss are not only able to reproduce the original sample, but are able to produce new sequences which conform to the structure but were not included in the original sample. One of the first studies illustrating this phenomenon was an experiment by Esper (1925). The stimulus materials in his experiment can be represented in a matrix, with the columns being different nonsense shapes and the rows different colors. The form of a stimulus from a particular cell was determined by the row and column it represented. Each of the nonsense forms and each of the colors were represented by a nonsense syllable in the response system. When a stimulus was presented the appropriate response was the nonsense syllable associated with its color followed by the nonsense syllable associated with its shape. The formal rules of the system, then, consisted of 1) vocabulary rules stating which nonsense syllables could appear; 2) a set of semantic rules stating the relationship between the nonsense syllables and what they referred to (i. e., color or shape); and 3) a syntactic rule which specified that the nonsense syllable referring to color must precede the one referring to shape. The Ss were not informed of the system inherent in the materials but were simply presented with stimuli and learned to give the appropriate double-syllable nonsense word. During the original learning the Ss were exposed to the materials from all of the cells in the matrix except three. After the original learning was completed a test trial was given in which not only the original learning materials were presented but the three stimuli the Ss had never seen were also presented. The Ss not only did as well on the three new stimuli as they did on the old stimuli, but they seemed not to be aware the stimuli were new.

While the results of the Esper (1925) experiment seem to suggest that the acquisition of structure cannot be accounted for in terms of the acquisition of simple habits through continuous reinforcement, and that some generalization mechanism is operating as well, the semantic rules that he used would allow for considerable transfer from the Ss' previous experience with English. An experiment by Kirk Smith (1964) is an illustration of how the kind of generalization apparent in the Esper experiment can occur when semantic rules are absent and there is no possibility of transfer from prior language habits. Smith's experiment also can be represented as a matrix. The matrix that Smith used is similar to that presented in Table 2. The system differed from that of Esper (1925) in that there were no specific stimuli paired with responses, and, consequently, there were no semantic rules. The rules of the system were: 1) the only response elements that can occur are those in the margins of the matrix; 2) all responses consist of two elements; 3) the only letters which can occur as first elements in the responses are those in

TABLE 2

	B	Z	G	K	V
X	X				
L		X			
F			X		
W				X	
N					X

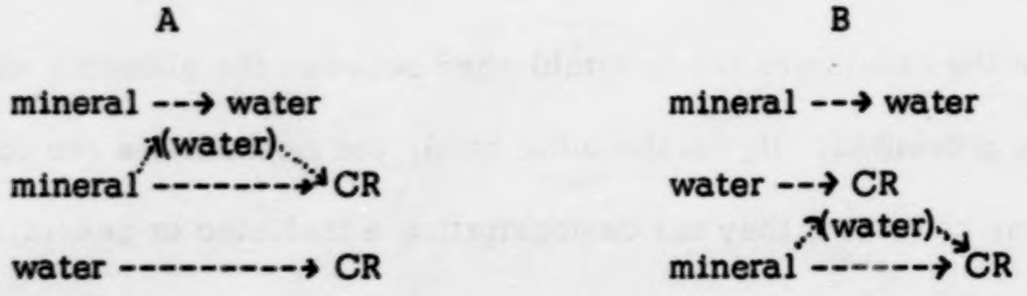
the row margins; and 4) the only elements which can occur in the second position are the letters in the column margins. The set of rules will generate 25 possible responses and the procedure was to read a set of the responses to the Ss and ask them at the end to recall those they had heard. In the list read to the Ss, however, the responses from some of the cells were not presented (e. g., those which are starred). Smith examined the Ss errors to check on the relative frequency that items from the starred cells occurred as errors as compared to responses which were wrong because they violated one or more of the rules of the system. His error data indicate quite clearly that Ss were learning something about the starred cells as they were learning the unstarred cells.

The results of the Smith (1964) and Esper (1925) experiments indicate there are implications of rule-conforming behavior in addition to the unequal transitional probabilities between the elements. This "something else" is an increase in the transitional probabilities for some unreinforced cells when the transitional probabilities of other cells are increased through specific reinforcement.

Jenkins and Palermo (1964) give a rather extensive and detailed analysis of the Esper (1925) experiment and demonstrate how the systematic generalization phenomenon apparent in the experiment can be accounted for in terms of mediation mechanisms, and they give a fairly detailed specification of the functional characteristics of the necessary mechanisms. In addition, there is a rather sizeable body of data demonstrating that these mediation mechanisms will account for the kind of generalized responding evident in the Esper (1925) and Smith (1964) experiments (Jenkins and Palermo, 1964; Horton and Kjeldergaard, 1962).

In attempting to give a psychological account of knowing-a-rule, whether these mediation mechanisms operate during original learning to establish habits between the elements which are not being reinforced, or whether the mediation mechanisms operate at the time of testing when the person is producing the previously unreinforced response, is a rather critical issue. If the mediation mechanisms operate during original learning, then the habits are being established between the elements which are not being presented. If, on the other hand, the mechanisms are operating at the time of testing they are demonstrating a mediated or generalized association and probably there has been no real association established. Viewing this distinction in terms of a transitional probability matrix, if the mediation mechanisms operate during original learning then equal transitional probabilities in two cells will always reflect equal underlying habits, even in the case where one of the cells has had its habit strength established through direct reinforcement and the other is a generalized habit. If the mediation mechanisms operate at the time of producing behavior (i. e., the test stage), and not during original learning, then a cell which has not been directly reinforced will have a zero habit strength, even though the transitional probability could be quite high. Relative to the problem of giving a functional characterization of knowing-a-rule, if the mediation mechanisms operate during original learning then these mechanisms must be included only in an account of how the behavior was acquired. If, on the other hand, the mediation mechanisms operate at the time of behaving then they must be included in any functional characterization of knowing-a-rule.

Razran (1949) studied the question of when mediated generalization occurs in an experiment which involved conditioning salivary responses to word stimuli. The mediation paradigms he used can be diagrammed as follows:



He assumed a uni-directional association between *mineral* and *water* (i. e., *mineral* elicits *water* but *water* does not elicit *mineral*). In addition, he assumed that the association from *mineral* to *water* was sufficiently strong that if the *S*s saw the stimulus *mineral* it would implicitly, if not explicitly, arouse the response *water*. In the first paradigm (A) he conditioned a response to the stimulus *mineral* but he assumed that *mineral* would elicit the response *water* implicitly and, consequently, the response would be conditioned to both *mineral* and *water*. On the test stage, when *water* is presented it should elicit the conditioned response. In this paradigm the mediating mechanism or event occurred only during the original learning. In paradigm B the response is conditioned to the stimulus *water*, but *water* does not elicit *mineral* and, consequently, no *mineral*-*CR* association or habit should be formed. On the test stage, however, when *mineral* is presented, it is assumed to elicit an implicit *water* which in turn elicits the conditioned response, i. e., the mediation occurs at the time of behavior and not during original learning. Razran's point was that if paradigm A elicited a greater generalized response than paradigm B it would suggest that mediated generalization is an original-learning phenomenon; whereas if paradigm B elicited a greater generalized response than paradigm A it would suggest that mediated generalization is a test-stage phenomenon. His results indicate paradigm B yielded greater mediated generalization.

Razran illustrated the same situation using another phenomenon. The paradigm in this experiment can be diagrammed as follows:

$$\begin{array}{lll}
\text{OL 1} & \text{Terrier} \dashrightarrow \text{Dog} \\
\text{OL 2} & \text{Dog} \dashrightarrow \text{Animal} \\
\text{OL 3} & \text{Dog} \overset{(\text{Animal})}{\dashrightarrow} \text{CR} \\
\hline
\text{TS 1} & \text{Terrier} \overset{(\text{Dog})}{\dashrightarrow} \text{CR} \\
 & \qquad \text{or} \\
\text{TS 2} & \text{Animal} \dashrightarrow \text{CR}
\end{array}$$

The word *terrier* tends to elicit *dog* as a response in word-association studies (OL 1), and *dog* tends to elicit *animal* as a response (OL 2). *Animal*, however, does not elicit *dog* nor does *dog* elicit *terrier*. Razran assumed that the first two original learning (OL) stages in the diagram were already established from the *Ss*' previous experience with their language. In the third OL stage he conditioned a response to *dog* and assumed *dog* would implicitly elicit *animal* as a response and thus allow some habit strength to exist between *animal* and the response. On the test stage (TS) he presented both *animal* (TS_1) and *terrier* (TS_2) as stimuli and measured the degree to which the response occurred to these two stimuli. If mediated generalization occurred to *terrier* it would indicate that the mediational mechanism was operating during the test stage; whereas if the conditioned response occurred more often to *animal* it would suggest that the mediation was occurring during OL. Again, Razran's data suggest that mediated generalization is a test-stage phenomenon, with *terrier* eliciting a greater generalized response than *animal*.

Recent data reported by Horton and Kjeldergaard (1962) does not support some of the specific findings of Razran. For example, they found paradigm A above yielding a somewhat stronger generalization effect than paradigm B. However, their data do support Razran to the extent that they demonstrate a significant test-stage generalization phenomenon when the possibility of OL generalization is minimized.

Another implication of the habit or transitional probability matrix conception of structure is that structure and frequency should be related. That is, if S̲s are presented with language sequences and asked to judge the degree to which they conform to the structure of language, the judgment should be determined primarily by how "ordinary" the sequences are or how frequently they have experienced similar sequences in the past. In an experiment by Maclay and Sleator (1959) S̲s were presented with a set of 36 sentences and word strings in which the meaningfulness, grammaticalness, and ordinariness of the sequences were varied independently of one another. Separate groups of S̲s were asked to judge the sentences along the three dimensions. The results of their experiment demonstrate that the S̲s' judgments of grammaticalness and ordinariness are quite independent of one another. In an unpublished replication of the Maclay and Sleator experiment, Hogan and Johnson (1964) had the same S̲s indicate degree of grammaticalness and ordinariness on a nine-point rating scale. The correlation between the two sets of judgments was only .30.

A final illustration of the inadequacy of a habit-matrix conception of knowing-a-rule is the recent data on the behavioral correlates of transformation rules. Chomsky's (1957) conception of transformational grammars suggests that simple, active, and declarative statements or kernel sentences are generated using phrase-structure rules and these are changed into other forms (e. g., questions, passives, negatives, etc.) by the application of one or more transformation rules. Miller (1962, a) has reported data indicating that the search time devoted to a set of sentences is correlated with the number of transformations necessary to understand the sentence. In addition, Gough (1964) has shown that if Ss are shown a sentence and later shown a picture, the latency in indicating whether the picture conformed to the sentence is related to number of transformation rules applied in generating the sentence. These two experiments, then, seem to suggest that the number of rules applied in generating a redundant sequence also has behavioral implications.

In summary, it has been supposed that the facilitating effect of grammaticalness (or rule conformance) on learning results from the fact that the constraints imposed by the rules either reduce the amount to be learned, or they allow Ss to use previous learning, or both. While the data generally seem to support that hypothesis they do not explicate the way Ss use the constraints.

Lashley (1951) has provided a rather convincing argument that, considering the speed of language responses, the constraints cannot take the form of item-to-item or word-to-word associations. If that is the case, then a question arises concerning not only the nature of the constraints, but, also, the nature of the responses they influence. It is with these issues that the research to be reported is primarily concerned.

As a first step it seemed reasonable to suppose that one of the influences of grammatical rules is that of defining the meaningful response units of language. That is, the rules may define not only the constraints but the nature of the responses they influence. In addition, it seemed necessary to identify these response units before any serious attempt could be made at determining the nature of the constraints.

CHAPTER II

Methods and Procedures

The general procedure used for this project was to conduct an extended series of small experiments which manipulated single variables rather than one large-scale experiment. At the beginning of the project there were very little data available indicating the variables which affected the influence of rules on behavior, nor were there any techniques available for examining the influence of rules on behavior. Therefore, it was assumed that a succession of small experiments, each one capitalizing on the data provided in the previous studies, would avoid placing a large investment in ineffective methods and techniques.

Although the project was conducted as an extended series of small experiments, there were a number of basic procedures which were used in most of the experiments. The commonality in the methods and techniques resulted from both the intent to study rule-governed behavior in learning situations and the hypothesis that one of the major influences of rules on behavior is that of defining functional decision units. The latter point led to the supposition that when Ss are learning rule-defined sequences of any kind, those response elements which belong to the same functional decision units should be more dependent as recall events during learning than are elements from different functional decision units. Therefore, unless otherwise specified, the following methods and procedures were followed in the experiments to be reported.

Experimental Task

The basic task used throughout the project involved a paired-associate learning situation. The Ss were required to learn a response sequence of some kind (e. g., a sentence) as a response to a digit. A standard paired-associate anticipation procedure was used in which the stimulus was presented for a short period of time during which the Ss were supposed to anticipate the response. The stimulus and response were then presented together and the Ss were to correct any errors they made. Then a new stimulus was presented for an anticipation interval followed by a presentation of its appropriate response, etc. Unless otherwise specified, a four second anticipation interval was used followed by a four second simultaneous presentation of the stimulus and response (i. e., a four second: four second rate). The Ss were required to learn several pairs which were presented in this manner and a trial consisted of a single presentation of all of the pairs the Ss were to learn. On each trial the order in which the individual pairs were presented was randomized and the trials continued until the Ss reached a criterion of either a fixed number of trials (a trials criterion) or they reached a specified level of performance on some particular trial (a learning criterion). The materials were presented to Ss on a Lafayette memory drum (Model 303 B). The pairs to be learned appeared in pica type on white paper.

The experiments were conducted in well-lighted and air-conditioned rooms which were seven feet wide and eight feet long. The rooms were furnished with a table for the memory drum and two chairs, one for the experimenter and one for the S. When the Ss arrived at the experimental room they were seated and given paired-associate learning instructions which described the learning task. The instructions also emphasized the importance of guessing during the anticipation interval. The latter point was made because there is a marked tendency for Ss who are learning a paired-associate task not to utter a response unless they are very certain it is right. The net effect is a large number of errors of omission during learning and very few errors of commission or partial responses. In that the technique to be used depended upon partial responses, it was felt that instructions which emphasized the importance of guessing were necessary. In addition, during the first few inter-trial intervals the Ss were reminded that they should guess during the anticipation interval, even when they were uncertain as to the correct response. These instructions seemed to be effective because throughout the series of experiments the proportion of errors of commission and the frequency of partial responses was in excess of those usually obtained in paired-associate situations.

Subjects

Except for two experiments the Ss used in the series of studies were introductory psychology students from the Ohio State University who participated as part of their course requirement. The population from which the samples are drawn varies between 1,000 and 2,000, depending upon the number of students enrolled in the course for that term. Within the population, females out-number males approximately 3 to 1 and that represents the approximate percentages of males and females in the samples for the individual experiments. However, early data indicated there was no significant sex difference in the phenomenon studied. In addition, the population consisted of approximately 60% freshmen, 25% sophomores, and the rest of the Ss equally divided between juniors and seniors. The Ss were obtained by posting a sign-up sheet near the room in which they have their psychology class. All Ss were native speakers of English, and they had never participated in an experiment similar to the one in which they engaged.

Scoring Techniques

The experimenter kept a verbatim record of all the Ss responses during the course of learning. The experimenter had before him a sheet of paper upon which were reproduced all of the responses the Ss should be making, in the order in which they should be made. The experimenter simply put a line through those words the S emitted correctly and noted any additional responses which were made at the appropriate point in the sequence. The Ss' responses were then scored for the probability of a transitional error between each succeeding pair of words within the sequence. A transitional error was defined as occurring when one of the responses in the sequence was wrong and the immediately preceding item was right. For each word-to-word transition the transitional error frequency was divided by the frequency that the word before the transition was correct to determine the probability of a transitional error (TEP). That is, the frequency of a transitional error was divided by the number of opportunities for a transitional error. A transitional error probability (TEP) was computed for each element-to-element transition within a response sequence. The number of transitional error probabilities, then, was the number of elements minus one. Each S learned several response sequences that were structurally identical and the transitional error probability (TEP) for any transition within the sequences was computed across all sequences and all trials for each S.

CHAPTER III

Pilot Experiments

Miller (1962, b) has suggested that messages which contain sequential redundancies can be coded more efficiently if segments are stored before the coding is attempted. That is, as the message is being received the S may be storing the items until a certain segment has been received. He then would take that segment out of storage and interpret it as the next segment is being received. Miller suggests that these segments probably correspond to phrases.

The point made by Miller seems to be that we do not handle structured language sequences in a word-to-word manner, but rather in somewhat larger chunks or decision units. It also may be the case that Ss use these larger decision units when producing language as well as when perceiving it. That is, Ss may make all of the decisions necessary to produce the words within a phrase prior to generating any of the words within it, rather than making the decision for each individual word just prior to emission. If Ss do make generative decisions which span several response items, then the occurrence of these response items should be more dependent upon the occurrence of other items within the same decision unit, than they are upon items from other decision units. This idea will be referred to as the unit hypothesis.

Viewing the hypothesis in somewhat more concrete terms it would be predicted that the conditional probability that an S will make an error on an item within a single decision unit, when the preceding element from the same unit was correct, should be relatively low as compared to the conditional probability of an error on the first element of a decision unit, given the last element of the preceding unit was correct. Therefore, if Ss learn sentences as responses in a paired-associate task and their response attempts are scored for transitional error probabilities (TEP) between words, the TEPs should be high on the transition between decision units and relatively low within decision units. The purpose of the first set of experiments was to explore the extent to which Ss used units larger than the word when learning sentences and to determine the extent to which the pattern of transitional error probabilities (TEPs) would detect these units. As a first approximation it was hypothesized that Ss would deal with sentences in phrase-sized units.

Experiment I

The first experiment was designed to explore the extent to which the number of phrases in a sentence would determine the difficulty of learning the sentence when the number of words and letters was controlled. It was hypothesized that learning difficulty should increase as the number of units or phrases increased.

Method

The learning procedure used in the experiment conformed to that outlined in Chapter II with the exception that the Ss learned the sentences to a criterion of one perfect trial. The sentences contained in Table 3 were used as responses.

TABLE 3

A

1. The tall boy saved the dying woman.
2. The strange man ate the fresh bread.
3. The angry lady broke the green dishes.
4. The young man found the wrong number.

B

5. The small child picked a blue flower.
6. The old friend gave a large party.
7. The hard work tires many young men.
8. The busy farmer likes the old horse.

C

9. The girls who live here are dancing.
10. The shoes with red bows were sold.
11. The fish from the lake are cooking.
12. The sound that I heard was singing.

D

13. The child who sings well was found.
14. The dogs with black tails are running.
15. The men who left early are listening.
16. The house across the street is burning.

The first eight sentences consist of two phrases and the structure of these sentences is represented in Fig. 1. Sentences 9 through 16 contain three phrases and their structure is represented by the tree diagram in Fig. 2. From these 16 sentences 2 lists were constructed with 8 sentences per list. The first list consisted of the first 4 sentences and sentences 9 - 12. The second list consisted of sentences 5 - 8 and 13 - 16. Fifteen Ss learned each list to a criterion of one perfect trial, and the Ss were assigned to the lists in alternating order as they appeared for the experiment.

Results

The Ss were scored for the mean trial upon which their four two-phrase sentences were correct for the first time as well as the mean trial upon which their four three-phrase sentences were correct for the first time. The significance of the difference was tested using a t-test for correlated means. The results were quite surprising in that they were exactly opposite to that which was predicted. The Ss learned the three-phrase sentences more rapidly than the two-phrase sentences. The Group 1 Ss learned their two-phrase sentences in 11.1 trials, whereas they learned their three-phrase sentences in only 9.1 trials. The Group 2 Ss learned their two-phrase sentences in 8.6 trials and they learned their three-phrase sentences in 6.5 trials. In both cases the difference was significant beyond the .05 level using a two-tailed test. Of the 30 Ss in the experiment, 25 learned their three-phrase sentences to the criterion more rapidly than their two-phrase sentences ($p < .05$).

One possible explanation for these results is that the Ss perceived the three-phrase sentences as being somewhat more interesting and adopted a strategy to learn the interesting sentences first. As a check on that hypothesis, all of the sentences used in the first experiment were reproduced on slips of paper 2" x 8 1/2", with one sentence per slip. Below each sentence was a nine-point rating scale. The slips of paper were then put into packets which consisted of all 16 sentences, one per slip. The packets were then distributed to the two groups of 40 introductory educational psychology students. Each student received two packets. On one packet the Ss were to indicate, on the scale, how interesting they thought the sentences were. On the other packet they were to indicate how grammatically complex they thought the sentences to be. One group made the interest judgment first and the other group made the complexity judgment first. The results indicate that while the Ss did perceive the three-phrase sentences as being significantly more complex than the two-phrase sentences, there was no difference in the interest value of the two types of sentences.

Although the rating scale data indicate that the sentences did not differ in interest value, the fact that they did judge the three-phrase sentences to be more complex does allow for the possibility that the Ss could have selected sentences to be learned on the basis of some complexity factor. The issue at stake in this consideration is whether the results of the first experiment indicate the three-phrase sentences are easier to learn or whether the Ss simply selected them to be learned first. A mixed-list design was used in the first experiment (i. e., two different kinds of responses in the same list), and any differences which appear between the two classes of response can be attributed to either a difference in learning ease or a difference in the extent to which the Ss prefer the responses. The latter case would not require there to be any difference in the ease of learning. Results of Postman and Riley (1957) indicate that differences in the rate at which two response classes are learned in mixed or heterogeneous lists can disappear when the lists are unmixed or homogeneous.

In the present experiment the three-phrase sentences of Group 1 and the two-phrase sentences of Group 2 were learned at about the same rate, even though they were sentences of different types. In addition, the two-phrase sentences of Group 1 and the three-phrase sentences of Group 2 were learned at extremely different rates. If the difference in the present experiment represents a selection strategy, rather than a learning-ease factor, then if the three-phrase sentences from Group 1 and the two-phrase sentences from Group 2 are put into the same list, the three-phrase sentences should be learned significantly more rapidly. If the two-phrase sentences learned by Group 1 and the three-phrase sentences learned by Group 2 are put into the same list the large difference obtained in experiment 1 should be narrowed considerably. If, however, the means for these four subgroups of sentences should approximate those obtained in experiment 1, then the experiment 1 results might be attributable to a learning-ease factor.

Experiment II

The second experiment was designed to explore the extent to which the results from experiment 1 should be attributed to a learning-ease factor or a selection strategy. In addition, the learning criterion in the first experiment was one perfect trial and, consequently, for some Ss, who learned the task very rapidly, there were data for only a very few trials. Therefore, TEPs could not be computed to test the unit hypothesis. Experiment 2 used a trials criterion rather than a learning criterion. In addition, in the first experiment there were only 15 Ss per group. The sample was doubled in the second experiment.

Method

Again the standard paired-associate learning task described in Chapter II was used. All Ss received 13 learning trials, irrespective of their level of performance. A 4-second: 4-second rate of presentation was used with a 4-second intertrial interval.

One group of 30 Ss learned the A and D sentences of Table 3 and the other group of 30 Ss learned the B and C sentences. The sample consisted of 60 Ss taken from the population described in Chapter II.

Results

The Ss were scored for the mean number of words given correctly across all trials for the two-phrase sentences, as well as the three-phrase sentences. The Group 1 Ss did learn the D sentences significantly more rapidly than the A sentences. The Group 2 Ss , however, did not learn the B and the C sentences at significantly different rates. In addition, however, the over-all rate of learning the three-phrase sentences was not significantly different from the rate for the two-phrase sentences. That is, while there was a significant difference for the Group 1 Ss the overall effect was not significant.

Because the Group 1 Ss did learn the two types of sentences at significantly different rates, when the Group 2 Ss did not, the results of the first experiment probably can be attributed to a learning-ease factor rather than a selection strategy. However, the failure to obtain an overall difference between the learning rates for the three-phrase sentences and the two-phrase sentences suggests that the differences obtained in this experiment, as well as in experiment 1, might be attributed to differences in the ease of learning the specific sentences used rather than a difference between two-phrase and three-phrase sentences.

In addition to scoring the Ss responses for the mean number of words correct, there were also scores for the conditional probability that each word would be wrong given the preceding word was correct. That is, the probability of a transitional error or TEP was computed for each of the six word-to-word transitions within the three-phrase sentences, as well as within the two-phrase sentences. These results are presented in Table 4.

If the Ss learned the sentences in terms of phrase-size decision units then there should have been a TEP spike on transition 3 for the two-phrase sentences and on transitions 2 and 5 for the three-phrase sentences. As can be seen by an inspection of Table 4, with the exception of the large TEP on

TABLE 4

Sentence Type	Transitions					
	1	2	3	4	5	6
Two-Phrase	.12	.05	.09	.06	.05	.02
Three-Phrase	.02	.13	.03	.02	.09	.01

transition 1 for the two-phrase sentences, the results seem to conform to that prediction. The Ss were scored for the mean TEP on transition 3 for the two-phrase sentences, as well as the mean of the other five transitions for those sentences. For the three-phrase sentences Ss were scored for the mean TEP on transitions two and five and the mean TEP on the other four transitions. The difference between the intra-phrase TEPs and the inter-phrase TEPs for both sentence types was tested using a t-test for matched pairs and in both cases the difference was significant beyond the .01 level. As predicted then, when Ss learned sentences in a paired-associate task, words within the same phrase were more dependent as recall events than were words from different phrases.

Discussion

There were a number of difficulties with the first two experiments and consequently, the results should be interpreted with some caution. First of all, the experimenters were inexperienced with the techniques being used and, consequently, some error could be introduced in that manner. Experiments 1 and 2 were performed by different experimenters and in both cases it was the first experiment in which they had engaged. Secondly, both experiments 1 and 2 involved a mixed or heterogeneous list design. While the results of experiment 2 seem to suggest that the differences in rate of learning the sentences cannot be attributed to that factor, the possibility of some kind of interaction between different sentence types in the same list cannot be completely ruled out. For example, if breaking the sentences up into smaller decision units represents a strategy the Ss adopt, then the fact that they have two different such strategies within the same list might result in confusion, and consequently, the TEP data may not be wholly representative of what the Ss are doing.

Although the above considerations should be kept in mind, it seems possible to conclude, at least tentatively, that for the sentences used in these experiments there is no overall difference in the ease with which the two-phrase and three-phrase sentences are learned. Secondly, the data from experiment 2 seem to suggest that Ss do learn phrases as relatively independent subunits when attempting to learn sentences.

Experiment III

Experiment 3 represented an attempt to capitalize on the information obtained in the first two experiments. As mentioned above, the fact that inexperienced experimenters were used for the first two experiments, as well as mixed lists, suggests that the data should be interpreted with some caution. These difficulties were eliminated in experiment 3 by the use of unmixed lists and an experimenter who was experienced.

Method

Again, the standard paired-associate learning procedure described in Chapter II was used. The sentences which the Ss learned were those in Table 3. Group 1 learned the A and B sentences, and Group 2 learned the C and D sentences (i. e., unmixed lists were used).

The Ss were 56 introductory psychology students who participated as part of their course requirement. Half of the Ss learned one list and the other half learned the other list. The Ss were assigned to lists in alternating order as they appeared for the experiment.

Results

Sentence learning rates. To test the difference between the groups in rate of learning the sentences, each S was scored for the total number of words given correctly across the 13 trials. The mean for Group 1 was 337, and the mean for Group 2 was 365, ($t = .82$, $df = 54$, $p > .40$). That result is consistent with the data from experiment 2 in demonstrating no significant difference in the rate at which the two sentence types were learned.

To check on the possibility that the Ss were integrating the sentences in serial order (i. e., from first word to last) they were scored for the frequency, over the 13 trials, that the first three words in the sentences were correct and the frequency that the last three words were correct. If there was any tendency to integrate the sentences in serial order, the first three words should be correct more often than the last three words. For both groups the first words were correct more oftan than the last three words, (Group 1: $t = 1.89$, $df = 27$, $p < .05$; Group 2: $t = 5.9$, $df = 27$, $p < .001$). The mean probability of an error for each of the seven word positions for the Group 1 sentences were: .519, .542, .522, .549, .551, .540, and .532. The respective means for Group 2 were: .444, .446, .518, .503, .509, .550, and .520. Although the first three words in the sentences were emitted more often than the last three words, these means indicate the tendency was not very great.

Pattern of transitional error probabilities (TEPs). The TEP for each of the transitions for each group is presented in Table 5 and Fig. 3 and Fig. 4.

TABLE 5

Sentence Type	Transitions					
	1	2	3	4	5	6
Two-Phrase	.11	.05	.12	.07	.03	.02
Three-Phrase	.02	.14	.03	.02	.12	.02

The TEP was computed for each of the transitions for each Ss (summing across all 8 sentences and 13 trials), and a Friedman two-way analysis of variance was performed. For both Group 1 and Group 2 the six TEPs were reliably different from one another. The X_r^2 for Group 1 was 62.01 ($df = 3$, $p < .001$) and for Group 2, X_r^2 was 62.90 ($df = 5$, $p < .001$).

In order to determine if the between-phrase TEPs were greater than the within-phrase TEPs, the Ss were assigned, as a score, the mean TEP on the between-phrase transitions for the sentences they learned and the mean TEP for the within-phrase transitions. For Group 1 the between-phrase transition score was the TEP on transition 3 and the within-phrase transition score was the mean TEP on the other five transitions. For Group 2 the between-phrase transition score was the mean of transitions 2 and 5 and the within-phrase score was the mean of the other four transitions. For both groups the difference was tested using a Wilcoxon matched-pairs signed-ranks test. For Group 1 the respective means were .120 and .059 ($z = 3.14$, $p = .0008$) and for Group 2 the means were .129 and .026 ($z = 4.62$, $p < .00003$). As in the previous experiment, the TEPs for the between-phrase transitions were greater than for the within-phrase transitions.

If the Ss learned the sentences in phrase-size units, then it seems clear that adjacent words from the same phrase should be more dependent as recall events than are the last word of one phrase and the first word of the next phrase. However, there is no reason for expecting differences between transitions within the same phrase. For both phrases in the sentences learned by Group 1, and the middle phrase of the sentences learned by Group 2, there is more than one within-phrase transition. A Wilcoxon test was used to determine whether the TEPs for different transitions within the same phrase were unequal.

For the middle phrase of the sentences learned by Group 2 $z = 1.50$ ($p = .1336$), and for the first phrase of the sentences learned by Group 1 $z = 2.34$ ($p = .0192$). A Friedman two-way analysis of variance was used for the second phrase of the sentences learned by Group 1, and $X_r^2 = 84.9$ ($df = 2$, $p < .001$). There were reliable differences, then, between TEPs for transitions from the same phrase.

The diagrams of the two types of sentence used in this experiment (i. e., Fig. 1 and 2) can be viewed as representing the immediate constituent (IC) structure of the sentences. Each of the nodes or circles in the diagrams represents a construction and the lines leading from a node indicate the two constituents into which it is to be divided. In a sense, then, the nodes represent the point in the IC analysis of the sentence where a division is to be made between the last word of the first constituent and the first word of the last constituent. At some point or level in the IC analysis every word is divided from its adjoining words by nodes and, within limits, it seems possible to rank order the transitions in terms of the "level" at which the node for that transition occurs. For example, for the sentences learned by Group 2 the first division occurs between the subject and the predicate at transition 5. The second division occurs between the initial noun phrase and the imbedded phrase at transition 2. The third division occurs at transition 3, and finally transitions 1, 4, and 6 are tied for fourth, fifth, and sixth. The transitions for the sentences learned by Group 1 can be rank ordered in the same manner.

If Fig. 1 is compared with Fig. 3 and Fig. 2 is compared with Fig. 4, it can be seen that the TEP for a transition is related to the level in the IC analysis that that transition represents a constituent transition. In order to determine the degree to which the obtained pattern of TEPs is related to the linguistic structure of the sentences the transitions can be rank ordered both in terms of TEPs and IC level, and a correlation computed. For Group 1 the Kendall rank correlation (tau) between IC level and TEP was .64 ($p < .068$) and for Group 2 tau was .75 ($p < .068$). There does appear to be a relationship between IC level of the transition and its TEP.

The results of experiment 3 support those of experiment 2 in demonstrating no difference in the rates at which the two types of sentence were learned. Also, the tendency for between-phrase TEPs to be greater than within-phrase TEPs was replicated in experiment 3. In addition to confirming the results of experiment 2, the results of experiment 3 appear to indicate that the TEPs reflect not only the phrase units but the entire immediate-constituent structure of the sentences.

Experiment IV

The results of the previous experiment seem to suggest that one contribution of grammar is that of defining functional units which *S*s use in learning sentences. Glanzer (1962) had *S*s learn, as paired-associate responses, two nonsense syllables separated by either a function word (e. g., fud and caz) or the two nonsense syllables separated by a content word (e. g., fud house caz). His results indicate that *S*s learned the pairs separated by a function word more rapidly than those separated by a content word. He suggested that separating two nonsense syllables by a function word results in something approximating a phrase unit in language. He then suggests that the obtained difference in learning may be the result of the differential ease of learning units versus non-units. Experiment 4 was a further exploration of that hypothesis.

The experiment by Miller and Selfridge (1950) was interpreted as suggesting that the ease with which a sequence of words can be learned is related to the extent to which the sequence conforms to that which can occur in English, even though they do not represent a sentence. The question to be explored in experiment 4 concerns the extent to which there is a contribution to learning ease made by having the material conform to grammatical units, over and above the contribution made by having the sequence conform to a grammatical sequence of words. For example, The father of my friend gave us a ride and The friend of my father gave us a ride are equally probable sentences in English. In the first sentence there is a grammatical segment, father of my, which does not conform to a phrase unit. In the second sentence there is an equally grammatical sentence segment, of my father, which does conform to a grammatical unit. In addition, both segments consist of the same words. If the influence, on learning, of having the material conform to acceptable sequences in English is only that of preserving short-range associations, then these two sentence segments should be learned at exactly the same rate. However, if there is a contribution made by conforming to the unit structure of language, over and above conforming to the acceptability and probability of the sequence, then the segment of my father should be learned more rapidly than the segment father of my.

Method

Again, the standard paired-associate procedure described in Chapter II was used. The digits 1 to 8 were stimuli and sentence segments were the responses. The Ss learned the task to a criterion of 12 trials.

The phrases used as responses in the experiment were the following: 1) in her book; 2) from the train; 3) for their party; 4) with our friends; 5) of that story; 6) across his fence; 7) under your plate; 8) on my shoe; 9) found the wrong man; 10) gave the poor child; 11) pushed the young girl; 12) saved the dying boy; 13) paid the tall lady; 14) ate the small fish; 15) likes the old farmer; 16) heard the busy woman. Each of the above phrases represents an acceptable sequence of words in English (i. e., they can occur in grammatically acceptable English sentences). In addition, if each of the terminal nouns in the phrases are placed at the beginning of the phrase, rather than at the end, (i. e., book in her as opposed to in her book) the result is an equally acceptable sequence of words, but now the sequence consists of two partial phrase units rather than one complete unit.

The two major groups were used in the experiment. The first group learned phrases 1 to 8 and the second group learned phrases 9 to 16. Each of the two groups were divided into two subgroups. The first subgroup of Group 1 learned phrases 1 to 4 as they are presented above, and phrases 5 to 8 with the noun at the beginning of the sequence rather than at the end. The other subgroup learned phrases 5 to 8 as they are presented above, and sequences 1 to 4 with the noun at the beginning rather than at the end of the sequence. Likewise, the first subgroup of Group 2 learned phrases 9 to 12 as they are presented above and 13 to 16 with the noun at the beginning rather than at the end of the sequence. The second subgroup of Group 2 learned phrases 13 to 16 as they are shown above, and 9 to 12 with the noun at the beginning of the sequence.

The Ss were 80 introductory psychology students, with 20 of them learning each list. All of the Ss' responses were recorded verbatim and the Ss were scored for the total number of words given correctly in the phrases, as well as the total number of words given correctly for the non-phrases.

Results

For Group 1 the mean number of words correct for the phrases was 58.0 and the mean for the non-phrases was 45.2. The difference was significant beyond the .01 level ($t = 4.81$, $df = 20$, $P < .01$). For Group 2 the mean number of correct words for the phrase units was 69.0 and for the non-phrase units the mean was 60.2. The difference was significant beyond the .05 level ($t = 1.81$, $df = 20$). For both groups then, the word sequences which conformed to phrase units were learned significantly more rapidly than equally grammatical sequences which did not conform to phrase units. The results of the experiment suggest that there is a contribution to learning by grammar, over and above the fact that it tends to preserve the short-range associations that Ss have established in their previous experience with language. Specifically, the contribution seems to stem from the fact that the grammar defines the units with which Ss handle the material, and they either prefer, or find it easier, to handle the material which conforms to those units.

Discussion of the Pilot Experiments

The results of these four experiments seem to leave little doubt that Ss tend to handle structured language material in terms of functional subunits. The results of experiment 3 indicate that Ss have a striking tendency to recall phrases as units during the course of learning sentences. The results of experiment 4 indicate that Ss handled grammatical segments, which also conformed to functional units, in a way which is different from that which they use for equally grammatical segments but which do not conform to functional behavior units. The problem is to explain the TEP data with the hope that the explanation will account for the way in which Ss use grammatical constraints or rules in producing language behavior.

It was noted in experiment 3 that the prbability of a transitional error (TEP) was correlated with the level in an immediate constituent analysis that a word-to-word transition first represents a constituent transition. It might be supposed that the immediate constituent structure of the sentence might determine the order in which the Ss integrate the words bounding the various transitions. However, an examination of the TEPs obtained in experiment 3 shows two marked deviations from that predicted by the IC level metric. The first deviation occurred on transition 1 of the Fig. 1 sentences. The TEP on the transition from The to the first adjective in the sentence was considerably greater than would be predicted by the IC level metric. Replications of the experiment with other sentences indicate that the deviation on that transition probably was not a characteristic of the structure, but of the particular adjectives used in the second word position. In some of the replications the data were similar to that obtained in experiment 3. In other replications, however, a deviation sometimes occurred on the transition to the second adjective in the sentence, but not on the first, and sometimes there was no deviation at all. In that the sentences used in these replications were structurally identical to those used in experiment 3, and yet there was fluctuation in the TEPs on transition 1, suggests that the fluctuation should be attributable to something other than structure.

The second eviation from that predicted from the IC level metric was on the Fig. 2 sentences. The TEP on transition 2 was greater than that on transition 5, but the IC level metric predicts the opposite should occur. Replications of experiment 3, using other sentences, yielded exactly the same results as obtained in experiment 3, with the TEP on transition 2 being significantly greater than that on transition 5. In that the same effect occurs on transitions 2 and 5, irrespective of the words surrounding those transitions, it seems reasonable to conclude that the effect is the result of the structure and not the particular words used. That one consistent deviation on the Fig.

2 sentences suggests that the hypothesis upon which the IC level metric rests may not be completely adequate.

A second possible interpretation of the TEP data is that they simply reflect the unequal word-to-word habits that the Ss brought into the experiment. To explore that hypothesis a group of Ss learned the word pairs around the second, third, fifth, and sixth transitions of the Fig. 2 sentences in both paired-associate and serial lists. When these word pairs were imbedded in the sentences the word pairs surrounding the third transitions were learned or integrated more rapidly than the word pairs surrounding the second transition. Likewise, when learned within the sentences, the word pairs surrounding the sixth transition were learned or integrated more rapidly than the word pairs surrounding the fifth transition. If the TEP data in experiment 3 were the result of unequal pre-existing word-to-word habits, then when the above word pairs are put in serial and paired-associate lists the word pairs surrounding the third and sixth transitions should be learned significantly more rapidly than those around the second and fifth transitions. The result of these experiments, however, indicates no significant differences in the rates with which these word pairs were learned. As a matter of fact, what differences did occur were in the wrong direction (i. e., the pairs around transitions two and five were learned more rapidly than those around three and six). That result indicates that the TEP data obtained in experiment 3 cannot be attributed entirely to unequal pre-existing word-to-word habits.

A third possible interpretation of the TEP data is that while the between-phrase TEPs reflect the unit structure of the material, the unequal within-phrase TEPs might reflect a "snowball" effect in the integration process. That is, if Ss integrate the elements within subunits in serial order, it might be the case that the more elements which have already been integrated, the easier it is to add a new element to the sequence. For example, in a four-element sequence it might be easier to integrate the fourth element with the first three than it is to integrate the second with the first element. While the snowball hypothesis could account for the unequal within-phrase TEPs, it does not explain the unequal between-phrase TEPs.

A fourth interpretation of the data has been offered by Osgood (1964). He seems to suggest that the pattern of TEPs might be the result of something like the accumulating context. While Osgood's accumulating context hypothesis seems consistent with the data, it would be difficult to quantify the hypothesis so that specific predictions could be made.

A fifth possible interpretation of the data is similar to Miller's (1956) ideas regarding the recall of recoded response chains. He suggested that binary digits may be recoded into pairs with a new symbol representing the pair (e. g., 00 = 1; 01 = 2; 10 = 3; 11 = 4). A sequence of binary digits could be recalled by retaining the sequence of encoding symbols and decoding each, as it comes up, into its respective pair of binary digits. The recoding would reduce the number of elements to be retained. The recall of a sentence might be retained in a similar fashion. Fig. 5 is a graphic illustration of how such a procedure might occur. An environmental signal (stimulus) arouses Σ' which is some kind of encoding unit representing the whole sentence. It is then decoded into subject (*S*) and predicate (*P*). *P* is stored in the subjects' short-term memory (STM) while *S* is decoded into article (*T*) and modified noun (*MN*). *MN* is stored in STM while *T* is decoded into *The*. Then the subject retrieves *MN* from his STM and decodes it into adjective (*A*) and noun (*N*). He then retrieves *A* from STM and decodes it into *tall*. *N* is then retrieved and decoded into *boy*, etc. The general procedure, then, is to decode each succeeding encoding unit into its constituent units, storing the second in STM while the first is further decoded. When a word is generated the subject retrieves from his STM the most recently stored item and decodes it in the same manner. This hypothesis predicts that *S*s generate sentences by proceeding through a system of decoding operations which decode high order encoding units into lower order encoding units. Yngve (1960) has suggested a similar idea. It will be referred to as the decoding operation hypothesis.

CHAPTER IV

In the previous chapter, five hypotheses were discussed in an effort to account for the TEP data obtained in the pilot experiments. Of these five hypotheses, two were shown to be inconsistent with the data. The IC level hypothesis made predictions about the TEP pattern for the Fig. 2 sentences which were not supported by the data. The hypothesis based on the assumption of differential pre-existing habits was contra-indicated by the fact that the learning ease of word pairs taken from the sentences was not related to the TEP data for the same word pairs when they were embedded in sentences. The three remaining hypotheses were: 1) the snowball hypothesis; 2) the accumulating context hypothesis; and 3) the decoding operation hypothesis. Of these three hypotheses, the first to be examined was the snowball hypothesis.

Basically, the hypothesis assumes that the Ss treated the phrases as independent units during learning(hence, the between-phrase TEP spikes). But, it assumes the unequal within-unit or phrase TEPs resulted from the fact that the Ss integrated the units in serial order, and the more elements which had previously been integrated, the easier it was to add a new element. That should result in a left-to-right decrease in the TEPs within a phrase. If the snowball hypothesis is correct, then if Ss learn sequences which are structured into subunits consisting of several elements, but the elements within a subunit are not structured, the same pattern of decreasing TEPs within a subunit should occur as occurred in the sentences. If, on the other hand, the decreasing within-phrase TEPs resulted from the within-phrase structure, the TEPs within unstructured subunits should be low and equal.

Pilot data indicated that if Ss learned random sequences of letters as paired-associate responses and the sequences were broken into segments by leaving blanks between certain letters, Ss tended to use the blank to break the sequence into subunits. The subunits, however, were not broken into sub-subunits as are the phrase-size subunits of language. Therefore, while the sequences of random letters were structured into subunits, the subunits were not structured.

To test the snowball hypothesis, Ss learned, as paired-associate responses, two seven-letter sequences. The sequences were typed on the presentation tape with all the letters adjacent to one another except for spaces at either the third transition or the second and fifth transitions (e. g., SBJ FXKM vs. SB JFX KM). The unit structure of the letter sequences was the same as the phrase structure of the sentences used in the pilot experiments reported in Chapter III. If the snowball hypothesis is correct, then the same pattern of decreasing within-subunit TEPs that occurred for the sentences should occur when the Ss learn letter sequences.

Experiment V

Method

The standard paired-associate procedure described in Chapter II was used for this experiment. The seven-letter sequences were constructed by using the Underwood and Schulz (1960) tables for the associative frequencies of letter pairs. All the adjacent letters within the sequences had associative frequencies of 0 or 1. Single digits were used as stimuli. The digits and letter sequences used were: 1) SBJHFZC; 2) MKXVGWP; 3) TCXFZWP; 4) NGVHSBJ; 5) DYJHQGW; and 6) BVPKXFZ. Each S learned two pairs of digits-and-letter sequences, and no letter appeared more than once in the sequences learned by an S. Group 1 learned sequences 1 and 2; Group 2 learned sequences 3 and 4; and Group 3 learned sequences 5 and 6.

Each Group consisted of two subgroups. On presentation one subgroup had all letters in the response typed adjacently except for a blank space between the third and fourth letters (e. g., DYJ HQGW). The other subgroup had a blank between the second and third and between the fifth and sixth letters (e. g., DY JHQ GW). The sequences are referred to as two-unit and three-unit sequences, respectively.

The Ss were 60 introductory psychology students. There were 10 Ss in each of the six subgroups, and all of their responses were recorded verbatim. The Ss response protocols were then scored for TEPs.

Results

The TEP data for each transition for both major subgroups is given in Table 6. In each case the hypothesis that the six TEPs were equal was tested using a Friedman two-way analysis of variance. In both cases the hypothesis was rejected at the .01 level.

TABLE 6

Sentence Type	Transitions					
	1	2	3	4	5	6
Two-Unit	.0992	.1086	.3135	.2089	.1914	.1919
Three-Unit	.0704	.2989	.1695	.1531	.2782	.0495

If the Ss used the blank spaces to define subunits in the sequences, then the TEPs for those transitions should be significantly greater than for intra-subunit transitions. Each S was scored for his mean interunit TEP, as well as his mean intraunit TEP. The difference was tested using a Wilcoxon signed-ranks matched-pairs test. For the two-unit sequences the mean intraunit TEP was .16, and the mean interunit TEP was .35 ($z = 3.6$, $p < .0002$). The respective means for the three-unit sequences were .11 and .29 ($z = 4.1$, $p < .0003$). In both cases a one-tailed test was used. The TEPs within units were signifcantly lower than the TEPs between units, suggesting that the Ss did use the spaces to break the sequences into subunits.

The only subunits which could be tested for intraunit TEP equality were the two subunits of the two-unit sequences and the middle unit of the three-unit sequences. The other two subunits had only one intraunit transition. In all three cases the null hypothesis could not be rejected. For the first unit of the two-unit sequences $z = .68$ ($p < .25$) and for the second unit $X_r^2 = 3.5$, $p < .10$). For the three-unit sequences $z = .01$ ($p = .50$). Contrary to the expectation from the snowball hypothesis, the within-unit TEPs were not significantly different from one another. While it is true that it is dangerous to draw conclusions from failure to reject the null hypothesis, the data do seem to suggest that the conditions which result in a significant difference between interunit and intraunit TEPs do not yield significant differences between the within-unit TEPs when structure is absent.

Although the TEPs within a single subunit did not differ from one another, there did appear to be differences between the subunits in terms of the mean within-unit TEP. For example, for the two-unit sequences the TEPs within the first unit were about equal, as were those for the second unit; but there seemed to be a difference between the mean of the first unit and the mean of the second unit. The same seemed to be true for the three-unit sequences. To examine these differences, each S was scored for the mean intraunit TEP for each unit. For the three-unit sequences, the means were .07, .16, and .05. ($X_r^2 = 1.63$, $p < .01$). If the magnitude of the within-unit TEP was related to the size of the unit, then the first and last subunit should be approximately equal, but both should be significantly less than the middle subunit. The first and last subunits were not significantly different from one another ($p > .16$), but they were both different from the middle subunit beyond the .01 level. For the two-unit sequences, the respective means for the within-unit TEPs were .10 and .20 ($z = 2.73$, $p < .01$). Again, the within-unit TEPs were different from one another, and the difference seemed to be a function of the size of the subunit.

Discussion

The results of experiment 5, when compared with those of experiment 3, seem difficult to reconcile with the snowball hypothesis. The snowball hypothesis clearly predicts that, in the present experiment, the conditions which result in a significant TEP spike on the inter-unit transition should also result in significant variation within the subunits. However, the mean range of variation within a subunit was only slightly more than .01. While it is dangerous to draw conclusions from a failure to reject the null hypothesis, the absence of variation in the within-unit TEPs is impressive. It would seem difficult to account for both experiments 3 and 5 using the snowball hypothesis.

Experiment VI

The next experiment was designed to explore the accumulating context explanation for the TEP data obtained in experiment 3. The hypothesis would assume that within each phrase there was an increasing semantic determination, or associative arousal, of succeeding words as the number of preceding words within the same phrase increases. In addition, while there should be an increase in semantic determination within a phrase, there should be a marked drop in determination from one phrase to the next. That would result in a pattern of TEPs similar to those obtained in experiment 3.

The accumulating-context hypothesis assumes there is an interaction of the influence of semantics and syntax on behavior. Such an interaction has inductive support from several studies. Johnson (1961) had S̲s learn sentences which contained nonsense words. The nonsense words appeared either in a noun position or an adjective position. The nonsense words appearing in the adjective position always had one characteristic ending (e. g., ...pod) and those appearing in the noun position had another characteristic ending (e. g., ...lef). After learning the sentences, the S̲s were presented with a new set of sentences which contained a blank in either the adjective or noun position, and they were given a choice between a lef nonsense word or a pod nonsense word to fit into the blank. To perform adequatly on the test stage, the S̲s would have had to learn the connection between the endings and the part of speech during the original learning. Johnson's data indicate that the S̲s were able to solve this task only if they had had some previous experience associating meanings with the nonsense words.

Marks and Miller (1964) had S̲s learn four kinds of sentences. The first consisted of sentences which were both meaningful and grammatical. The second group was constructed by taking the first group of sentences and scrambling the words within the sentences in anagram fashion. These they called meaningful but not grammatical. The third group of sentences were grammatical but meaningless, and the fourth group of sentences were constructed by scrambling the third group in anagram fashion. The fourth group of sentences were neither meaningful nor grammatical. The learning data collected with these sentences indicate that the fourth group of sentences had the poorest rate of learning and the first group had the best rate of learning, with the second and third groups being about equal and falling half-way between the first and fourth groups. They conclude that semantics and syntax contribute equally to learning, and that the effects are combined the produce the relatively rapid rate of learning for sentences which are both grammatical and meaningful.

A final study illustrating the interaction of the effects of semantics and syntax on behavior is one by Prentice (1964). In her study, children were to learn nonsense words as responses to either pictures or sentences with a word missing. She assumed that the picture stimuli would define the semantic meaning of the nonsense words, and the sentences with the blank would identify the nonsense words' appropriate part-of-speech or form class. After the original learning, all Ss received two tests. For the semantics test, the Ss were asked to indicate which nonsense word had approximately the same meaning as a presented English word. On the grammatical test, the children were asked to use the nonsense words in some new sentences which were presented. In the latter test, the children could only solve the task if they had formed an association between the nonsense words and their appropriate part-of-speech during original learning. Although the group which had semantics training did best on the grammar test, both groups performed at a significantly better than chance level on the other test as well. That is, the group which had semantics training was able to use the nonsense words as their appropriate part-of-speech, and the syntax group was able to use their knowledge of the parts-of-speech to identify the meaning of the words.

The above experiments seem to suggest that reference or meaning can influence Ss behavior on tasks that appear to be tests of their grammatical knowledge. The results of experiment 3 could be explained in the same terms by assuming that the pattern of semantic determination or associative arousal is correlated with the immediate-constituent stucture of the sentences. On intuitive grounds, it does seem that such a correspondence or correlation does exist. There is a question, then, as to which of the two sources the TEP data should be attributed.

Semantic determination must exercise its effect through some kind of associative arousal. The data reported above, however, indicate that the TEP data cannot be attributed to differential pre-existing word-to-word habits. At least, the difficulty of associating words together when they are taken out of the sentence context is not related to the TEP between the words when they are within the sentence. It is possible, however, that a pre-existing association could exist between a word in the sentence and all of the preceding words, (i. e., the context). Experiment 6 was an attempt to determine if the pattern of associative arousal within a response sequence could influence the TEP pattern when other sources of structure were absent. The Ss learned, as paired-associated responses, sequences of seven words. The sequences were constructed using the Russel and Jenkins (1954) norms for the Kent-Rosanoff word association test. Within each of the sequences the first, fourth, and sixth transitions were spanned by relatively strong pre-existing associations, while the second and fifth transitions were not. The intent was to construct a unit structure similar to the phrase structure of the Fig. 2 sentences.

Method

The Ss learned six pairs consisting of a single digit as a stimulus and sequences of seven words as responses. The standard paired-associate procedure described in Chapter II was used for the experiment. The Ss received 13 learning trials and the materials were presented at a 4-second:4-second rate.

The sequences which the Ss learned were: 1) afraid fear stem flower smell stomach food; 2) bath clean fruit apple red mountain hill; 3) city town dream sleep bed mutton lamb; 4) doctor nurse cottage house home lion tiger; 5) eagle bird ocean water drink religion church; 6) foot shoe justice peace war needle thread. As can be seen, except for transitions 2 and 5, there is a fairly strong associative connection between adjacent words within each of these sequences.

The Ss for the experiment were 24 introductory psychology students. The Ss' responses were recorded verbatim, and TEPs were computed following the procedure described in Chapter II.

Results

The TEPs for each of the six transitions within the sequences were: 1) .094, 2) .489, 3) .214, 4) .375, 5) .482, 6) .107. As can be seen, the TEPs on transitions 2 and 5 were approximately equal and greater than any other transition within the sequence. To determine whether the six TEPs were different from one another, a Friedman two-way analysis of variance was performed. The differences were reliable at the .001 level (X_r^2 = 86.44, df = 5). To determine whether the strength of the pre-existing association influenced the pattern of TEPs each *S* was scored for the mean TEP on the two transitions with low associate strength spanning them, and the mean TEP on the four transitions with a high associative strength spanning them. The means were all considerably larger than zero, and a t-test for correlated means was performed. The difference was significant beyond the .001 level (t = 13.07, df = 23). To further test the strength of the effect, the significance of the difference between transitions 4 and 5 was tested. Transition 4 had the highest TEP of those transitions which had a high associative strength spanning them, and transition 5 had the lowest TEP of the two transitions which had a low strength spanning them. The difference was tested using a Wilcoxon and t = 40 (N = 24, $p < .005$). Even the smallest difference between high and low associative strength transitions reached an acceptable level of significance. While the data reported above indicate that the TEP data from experiment 3 cannot be accounted for in terms of pre-existing word-to-word habits, the results of experiment 6 indicate it is possible for such associations to influence the TEP pattern when other sources of structure are absent. Now the question to be answered is whether grammatical structure can influence the TEP pattern when the associative structure is eliminated.

Experiment VII

The accumulating-context hypothesis (semantic determination) assumes that the preceding context arouses succeeding words as a result of pre-existing associations, and that the strength of the arousal is a function of the amount of prior context. Experiment 6 indicated that Ss are able to use the assumed pre-existing associations when they are provided. That experiment, however, only indicated that it is possible for these associations to be used, and not that they are in fact used to produce the TEP pattern. However, if the pattern of TEPs is the result of semantic determination, then if a set of sentences is constructed which are grammatically correct, but meaningless to the extent that semantic determination is reduced to zero or near zero, then the pattern of TEPs obtained in experiment 3 should not be obtained. If the same, or a highly similar, TEP pattern is obtained with meaningless sentences, the data will provide a strong argument that Ss are using grammatical structure.

Method

The standard paired-associate procedure described in Chapter II was used for experiment 7. The Ss learned a six-item paired-associate list consisting of digits as stimuli and seven-word meaningless sentences as responses. The sentences conformed to the Fig. 2 structure. The sentences used were: 1) The falsity calling flat sleep sang white; 2) The possibility shining bent thirst waited truthfully; 3) The afternoon smiling green stories burned wet; 4) The humor swimming orange years froze carelessly; 5) The graciousness going tall centuries cooked closely; 6) The promise hearing slow staples thought round. The sentences were constructed such that the probability of each word following the previous word in natural English would be either zero or very close to zero, except for the transition from The to the second word of each sentence. It was assumed that there would be little or no semantic determination within these sentences.

All Ss received 15 learning trials at a 4-second: 4 second rate. The Ss were 20 introductory psychology students. Their responses were recorded verbatim, and their response protocols were scored for TEPs according to the method described in Chapter II.

Results

The mean TEPs for the six transitions were: 1) .1788, 2) .7343, 3) .6479, 4) .6076, 5) .7509, 6) .3878. To determine if the six TEPs were reliably different from one another, a Friedman two-way analysis of variance was performed, and $X_r^2 = 43.7$ (df = 5, $p < .001$). To determine the extent to which the elimination of semantic structure would change the pattern of TEPs obtained in experiment 3, the six TEPs obtained in each of the two experiments were rank ordered and a rank-order correlation (rho) computed. The correlation between the results of the two experiments was .89. To test the significance of the correlation, a correlation was computed for each of the Ss in experiment 7 between his pattern of TEPs and the overall pattern obtained in experiment 3 for the Fig. 2 sentences. Every S had a positive correlation ($p < .001$). In addition, a correlation was computed between the rank order of the TEPs obtained in experiment 7 and the IC metric described in the report of experiment 3. The correlation was .94 and, again, every S had a positive correlation ($p < .001$).

Discussion

The results of experiment 7 do not eliminate the possibility of some influence of semantic determination on the obtained pattern of TEPs. Certainly, the regularity of the pattern of TEPs obtained in experiment 7 does not even approximate that obtained in experiment 3. However, a comparison of the relative magnitudes of the TEPs obtained in experiment 7 with those obtained in experiment 3 seems to suggest clearly that the phrase structure of the sentences exerts an influence on the TEP pattern when the influence of semantic determination is reduced to a minimum.

Experiment VIII

Experiment 8 was an attempt to examine the viability of the IC level hypothesis, the semantic-determination hypothesis, and the decoding-operation hypothesis under conditions where the three hypotheses make different predictions for the same data. The materials used in the experiment were sentences which conformed to the structure represented in Fig. 6. The most interesting aspect of these sentences is the segment represented by the sequence so highly loved boy in Fig. 6. In that the nodes correlated with these three transitions become increasingly higher in the tree diagram from the second transition to the fourth transition, the IC level hypothesis would predict an increase in the TEPs across the three transitions. The contextual or semantic-determination hypothesis should predict a linear decrease in the TEPs across the three transitions, because of the apparent increase in contextual determination. At least, it would appear the contextual constraint across transitions 2, 3, and 4 should be similar to that across transitions 5, 6, and 7. The decoding-operation hypothesis predicts that the TEPs should be a function of the number of decoding operations necessary to generate the constituent following the transition. The number of predicted operations at each of the six transitions is given in Table 7. As can be seen, the decoding-operation hypothesis predicts that the TEPs on transition 2, 3, and 4 should be relatively low and equal; and the TEPs at transitions 5, 6, and 7 should decrease in that order. To summarize, then, all three hypotheses predict a linear decrease in the TEPs for transitions 5, 6, and 7; but given such a significant decrease for those transitions, the hypotheses differ as to what should occur on transi-

TABLE 7

	Transition Number						
	1	2	3	4	5	6	7
TEP	.94	.11	.12	.13	.57	.34	.06
Number of Operations	10	1	1	1	7	4	1

tions 2, 3, and 4. The IC-level hypothesis predicts a significant increase from transition 2 to transition 4; the semantic-determination hypothesis predicts a significant decrease in TEP from transition 2 to transition 4; and the decoding-operation hypothesis predicts transitions 2, 3 and 4 should not be significantly different from one another.

Method

The Ss learned eight paired-associates following the method described in Chapter II. The stimuli were the digits one through eight, and the responses were eight eight-word sentences which conformed to the structure illustrated by Fig. 6. The sentences which the Ss learned were: 1) A very kindly treated man walked two miles; 2) Those quite simply dressed ladies entered the house; 3) The somewhat frequently tired dog stayed at home; 4) That rather easily completed course started on time; 5) A too rapidly finished test failed to pass; 6) The most clearly delivered speech ended after dinner; 7) That so highly respected teacher talked to us; 8) Those least slowly traveled roads closed early today. The sentences were presented at a 4-second: 4-second rate, and all Ss received 15 learning trials.

The Ss were 25 introductory psychology students. Their response attempts were recorded verbatim, and their response protocols were scored for TEPs according to the method described in Chapter II.

Results

The results of experiment 8 are presented in Table 7, along with the number of decoding operations at each transition predicted by the decoding-operation hypothesis. A Friedman two-way analysis of variance was performed on the data and the differences between the seven means were reliable beyond the .001 level ($\chi_r^2 = 103.65$, df = 6). The three hypotheses make differential predictions for transitions 2, 3 and 4 given the predicted decrease in TEP across transitions 5, 6, and 7. Therefore, to determine if the basic condition of the experiment had been met, a Friedman two-way analysis of variance was performed on transitions 5, 6, and 7. The decrease in the TEPs was significant beyond the .001 level ($\chi_r^2 = 41.04$, df = 2).

The semantic-determination hypothesis leads to a predition that the TEPs for transitions 2, 3 and 4 should decrease in the same manner that the TEPs for transitions 5, 6, and 7 decrease. In that the TEPs increase slightly from transition 2 to transition 4 that expectation clearly was not supported by the data. The IC-level hypothesis leads to an expectation that the TEPs across transitions 2, 3, and 4 should show a linear increase which is similar to the linear decrease across transitions 5, 6, and 7. Clearly, that was not the case. To test the reliability of the differences between transitions 2, 3, and 4, a Friedman two-way analysis of variance was performed; and the differences were not significant ($\chi_r^2 = 0.96$, df = 2, $p > .50$). The failure to obtain significant differences between the TEPs at transitions 2, 3 and 4, when such differences are obtained across transitions 5, 6, and 7, is more consistent with the decoding-operation hypothesis than either the IC-level hypothesis or the semantic-determination hypothesis. The mean difference between succeeding points for transitions

2, 3, and 4 is only .01, whereas the mean difference for succeeding points across transitions 5, 6, and 7 is .26. The discrepancy occurred in that direction for every single *S*. The results, therefore, appear to offer further disconfirmation of both the IC-level hypothesis and the semantic-determination hypothesis.

General Discussion

The conceptual ground on which the preceding eight experiments were based might be termed the *unit hypothesis*. In its simplest form, it states that *Ss* produce language sequences in terms of subunits, and that there is a marked tendency for the *Ss* to produce or recall the subunits in an all-or-none manner. In addition, the probability of all-or-none recall of the unit should be a function of the size and/or difficulty of the unit. That led to the expectation that when *Ss* learn sentences as paired-associate responses, words that are within the same subunit should be more dependent recall events than are words from different subunits. The results so far reported seem to offer rather clear support for that hypothesis. The results, however, also indicate that the unit hypothesis must be elaborated.

At the end of Chapter III, five alternative elaborations of the unit hypothesis were offered as possible explanations for some of the finer details of the data obtained in experiment 3. The experiments described in the present chapter represented an attempt to examine the tenability of these hypotheses. The IC-level hypothesis was rejected because it could not account for the results of experiment 3 and experiment 8. The hypothesis based on the assumption of pre-existing word-to-word habits was rejected because the strength of these habits did not seem to be related to the TEP pattern. The snowball hypothesis was rejected because it failed to account for the results of experiment 5. The fact that the results of experiment 7 were quite similar to those obtained in experiment 3, when semantic determination was reduced to a minimum, suggests that the TEP pattern cannot be attributed entirely to the semantic constraints which exist within the sentences. Finally, the fact that the TEPs within left-branched sequences are not different from one another when the TEPs within similar, but right-branched, sequences are significantly different from one another, does offer some support for the decoding-operation hypothesis.

Before attempting to explore the empirical implications of the decoding-operation hypothesis, an attempt should be made to examine some of its details. To begin with, it assumes that S̲s proceed through the system of decoding operations outlined in Chapter III when they are attempting to generate a sentence. The S̲s decode each succeeding encoding unit into its constituents, storing all but the first constituent in short-term memory, and then decoding the first constituent in the same manner until a terminal event occurs. The S̲ then retrieves from his short-term memory the most recently stored item and decodes it in the same manner.

The encoding units, or items the S̲s are storing in their short-term memory, are assumed to be the sequences to which the unit hypothesis applies. Each time the S̲ retrieves an encoding unit from his short-term memory, he should tend to recall it in an all-or-none manner. The probability of a recall for any constituent unit should be a function of the size and/or difficulty of the unit.

If it is assumed that *S*s proceed through a system of decoding operations to generate a sentence, then the TEP on the transition to any unit should be a function of the number of decoding operations it contains or implies. Fig. 5 can be used to illustrate the procedure of counting decoding operations. When the *S* has decoded *T* into *The*, he retrieves *MN* from his short-term memory. Two decoding operations are involved in decoding *MN* into *A* and *N*. One decoding operation is required to decode *A* into *tall* and one is required to decode *N* into *boy*. In all, there are four decoding operations involved in completely decoding the constituent unit which follows the word *The*. After the *S*s has produced the word *tall*, he retrieves *N* from his short-term memory and in one operation decodes it into *boy*. Therefore, there is only on decoding operation involved in the constituent unit which starts with the word following *tall*. After producing *boy*, the *S* retrieves *P* from his short-term memory. There are 10 decoding operations necessary to completely decode *P*.

The decoding-operation hypothesis only describes the way in which *S*s use the structure of a sentence in the process of generating it. If the decoding-operation hypothesis and the unit hypothesis are combined, then a set of predictions can be made as to the relative magnitude of the TEPs at each of the transitions within a sentence when *S*s are learning it. That is, the decoding-operation hypothesis describes the units *S*s use, and the way in which they use them, and the unit hypothesis specifies the effect of such utilization on behavior during learning.

CHAPTER V

The experiments to be described in the present chapter were an attempt to test the viability of the decoding-operation hypothesis. One of the most interesting implications of the hypothesis is that it assumes Ss proceed through generation of a sentence by following a system of decoding operations, rather than executing word-to-word habits. It suggests that the relevant responses may be decoding operations and not the emitted words. If that is the case, then the associations Ss use should be between decoding operations and not the words themselves.

That implication can be tested by capitalizing on the fact that although all succeeding words within a sentence are adjacent verbal responses, the operations which produce them are not necessarily adjacent. For example, in the sentence provided in Fig. 5 both the word pairs tall-boy and boy-saved are instances of adjacent verbal responses. That is, the second word of the pair follows the first within the sentence. However, by examining the tree diagram presented in Fig. 5, it can be seen that while the operations producing tall and boy are adjacent, the operations producing boy and saved are not adjacent. That is, after the S decodes A into tall, he recovers N and decodes it into boy. After decoding boy, he does not recover V and decode it into saved. Rather, he recovers P from his short-term memory and decodes it into verb plus noun phrase. Only then does he decode verb into saved. Therefore, while the operations producing tall and boy are adjacent, those needed to produce boy and saved are not adjacent.

If, as assumed above, *S*s form and use associations between decoding operations as they are attempting to learn and recall the sentences, then *S* should attempt to form and use an association between the operations producing *tall* and *boy* but not between the operations producing *boy* and *saved*. Therefore, if a prior association is established between these two pairs before the *S*s attempt to learn the sentences, then there should be a significant reduction in the TEP for the *tall*-*boy* transition because they are being provided with an association they can use. However, no such reduction in TEP for the *boy*-*saved* transition should occur. On the other hand, if *S*s are forming and using associations between words as they are attempting to learn and recall the sentences, the reverse phenomenon should occur. That is, there should be a greater reduction on the *boy*-*saved* transition than on the *tall*-*boy* transition. That prediction stems from the fact that the influence of a learning experience (e. g., a trial) is inversely related to the amount to be learned. The typical negatively-accelerated learning curve is an illustration of this phenomenon. In that the prior data indicate that the TEP for the *boy*-*saved* transition will be greater than for the *tall*-*boy* transition, the influence of the prior learning experience should be greater on the *boy*-*saved* transition if *S*s use word-to-word habits. Therefore, there should be a greater influence from the prior learning experience on that transition. The following experiment was designed to test that hypothesis.

Experiment IX

Method

As in the previous studies, the Ss task was to learn an eight-item paired-associate list which consisted of eight seven-word sentences as responses and the digits one to eight as stimuli. The general procedure was that described in Chapter II. Before learning the sentences, however, the Ss learned a paired-associate list which formed an association between the second and third words (second transition) and/or the third and fourth words (third transitions). For each S the prior association was established for: 1) both the second and the third transitions for two sentences (high-high sentence class); 2) only the second transition for two sentences (high-low sentence class); 3) only the third transition for two sentences (low-high sentence class); and 4) neither transition for two sentences (low-low sentence class). Each of the eight sentences appeared in each of the four classes for an equal number of Ss (i. e., there were four original learning lists). Of the 24 words bounding the two transitions (i. e., three per sentence for eight sentences) only 14 occurred in the original paired-associate list. The 10 words which did not occur in the list were put in a familiarization list and Ss just pronounced those words after the original learning task was completed. After the paired-associate and familiarization trials, the Ss learned the sentences.

The sentences which the Ss learned were the following: 1) The tall man needs a bigger suit; 2) The new person did the wrong thing; 3) The large family wants a new house; 4) The fat woman took the large piece; 5) The last party was a great success; 6) The old lady told the funny story; 7) The young girl had a pleasant smile; 8) The two sisters liked the same boy.

The materials were presented at a 4-second: 4-second rate, and all Ss received 14 learning trials. The Ss were 80 college sophomores, and all of their responses were recorded verbatim. They were assigned to the various original paired-associate learning lists in alternating order as they appeared for the experiment.

Results

To determine the extent to which the adjective-noun (A-N) pairs and the noun-verb (N-V) pairs had been equally learned during original learning, each S was scored for the total number of A-N errors and the total number of N-V errors, summed across the 14 learning trials. The difference in favor of the A-N pairs was significant beyond the .01 level. To determine the extent of the difference in degree of learning of the two kinds of pairs, the mean percentage of correct responses during the last three trials was computed for each class; and the mean for the A-N pairs was .95 as opposed to .92 for the N-V pairs. Therefore, while there was a significant difference in rate of learning for the two kinds of pairs, they were quite close together in the degree of learning ultimately attained.

The TEPs for each transition, for each sentence class, and across all sentences are given in Table 8. The hypothesis that the six probabilities were equal was tested using a Friedman two-way analysis of variance, and the hypothesis was rejected at the .001 level ($X_r^2 = 167$, df = 5).

TABLE 8

Group	Transition Number					
	1	2	3	4	5	6
High-High	.0518	.0224	.0763	.0288	.0521	.0034
High-Low	.0519	.0371	.0866	.0244	.0694	.0042
Low-High	.0446	.0504	.0744	.0310	.0577	.0091
Low-Low	.0467	.0489	.0820	.0380	.0753	.0098
All classes	.0487	.0399	.0799	.0306	.0639	.0067

The decoding-operation hypothesis predicts that transition 3 should have a higher TEP than the others. The hypothesis was tested by scoring Ss for the mean TEP on transition three and the mean of the other five transitions. The respective means were .0799 and .0380. The difference was tested using a Wilcoxon signed-ranks matched-pair test. The difference was reliable beyond the .001 level. To determine if the two TEPs within the first phrase were unequal, a Wilcoxon test was used and $p = .0594$ ($z = 1.49$, $df = 79$). The differences within the second phrase were tested using a Friedman two-way analysis of variance and the hypothesis of equality was rejected at the .001 level ($X_r^2 = 74$, $df = 2$).

To evaluate the agreement between the obtained pattern of TEPs and that predicted by the decoding-operation hypothesis, both the predicted and obtained sets of TEPs were rank ordered and a rank-order correlation (rho) computed. The rho was .60. A separate rho was computed for each of the 80 Ss, and 72 of the rhos were positive, two were zero, and six were negative ($p < .001$).

To determine the influence of the previous association on the TEPs during learning, the TEP was computed for each of the two manipulated transitions for each S under both high and low previous associations. In both cases, the difference was tested using a Wilcoxon. For the second transition (A-N) the mean TEP under low prior association was .0496, and under a high prior association it was .0284 ($z = 2.374$, $df = 79$, $p < .0089$). For the third transition (N-V) the respective means were .0842 and .0760 ($z = .086$, $df = 79$, $p > .46$). As predicted by the decoding-operation hypothesis, the prior association resulted in a significant decrease in the TEP for the second transition but, under the same conditions, did not yield a significant change for the third transition.

In order to determine if the differential effect of the prior habit on the two transitions could be attributed to the different rates of original learning, correlations were computed between the number of right responses on the two classes of pairs during the original paired-associate learning and the degree of facilitation on the respective transitions during sentence learning. For the A-N transition, the correlation was -.13, and for the N-V transition the correlation was -.20. In addition, if the differential change on the two transitions was the result of differential degrees of original learning, then there should be a correlation between the differences between the number of right responses on the two kinds of pairs during original learning and the difference between the degree of facilitation from the prior habit during sentence learning. Again, however, there was not a significant positive correlation (rho = -.14). These low negative correlations would seem to suggest that the differential effect of the prior associations on the two transitions cannot be attributed to differences in degree of original learning.

The failure to obtain a significant influence from the prior association on the N-V transition, under conditions which did yield such a change for the A-N transition, lends inductive support to a model of language production which involves decoding high-order encoding units into subunits. While, to a degree, such an interpretation of the data entails the risks involved in making inferences from failure to reject the null hypothesis, there is a rather large difference in the relative mean change as a result of the prior association (i. e., $z = 2.374$ vs. $z = .086$).

Frankart (1964), in a partial replication of the above experiment, provided data which supported the result. She noted that in experiment 6, reported above, the TEPs within ungrammatical sequences of words were markedly influenced by prior associations, whereas the data of experiment 9 indicate that prior associations spanning phrase transitions do not influence the TEPs within grammatical sentences. She was concerned with the extent to which there might be a relationship between degree of grammaticalness and the extent to which prior associations can reduce TEPs. Her Ss rated a large number of word sequences, which varied from random sequences to grammatical sentences, as to their judged degree of grammaticalness. She then constructed four lists of eight sentences each. The lists differed in mean degree of grammaticalness from list one, which consisted of random sequences of words to list four which consisted of perfectly grammatical sentences. She used four groups of Ss, one per list. Half of the Ss in each group had a prior association established between a word pair within each of the sequences. The other half of the Ss were simply familiarized with the words, but the words from a pair from any one sequence were never presented together so no association was established. The Ss then learned the sequences as responses to digits following the procedure described in Chapter II. Her results indicate that the degree to which the prior habit reduced the TEP on the transition between the two words was a decreasing function of grammaticalness. That is, the more grammatical the sequence the smaller was the change from the prior habit. For the random sequences, the effect of the habit was sizeable, for the grammatical sentences there was no significant influence from the habit.

The results of both experiment 9 and the Frankart study support the decoding-operation hypothesis in that they demonstrate that Ss do not use word-to-word habits at all of the word-to-word transitions within sentences, and whether such a habit can be used for a given transition is predictable from the hypothesis. At the very minimum, the data suggest that an explanation for the TEP patterns which assumes that Ss assemble a series of word-to-word habits, will be inadequate. For example, the IC-level hypothesis assumes that Ss use the structure of sentences as a learning strategy which defines the order in which succeeding word pairs would be integrated. If that were true, the magnitude of the TEPs would be correlated with the point in the learning process that the transitions were integrated. Although that hypothesis would suggest that Ss delay the integration of certain word pairs in favor of the integration of other pairs, when the time came for an S to integrate a particular word pair, a prior habit should help him and thereby reduce the TEP for that transition. It would be quite difficult to reconcile that hypothesis with the data reported above.

Experiment X

Another implication of the decoding-operation hypothesis is that if decoding steps are occurring in the manner described, then they should occupy time. That is, as the number of hypothetical decoding operations increases at any particular point in a sentence, the time to perform them should also increase. Therefore, there should be an increase in the latency between the words bounding the transition. Data reported by Maclay and Osgood (1959) offer some support for that prediction. They tape-recorded a conference and then attempted to determine where hesitations of various kinds tended to occur in an Ss spontaneous speech. Their data indicate that approximately 50% of the hesitations tended to occur at phrase boundaries where the decoding-operation hypothesis predicts the greatest number of operations should occur.

If decoding operations occupy time, not only should there be greater interphrase latencies than intraphrase latencies, but the latency between the environmental signal which sets the occasion for generating a sentence and the first word of the sentence also should be a function of the number of decoding operations the *S* must make in order to generate the first word. For example, in the Fig. 2 sentence the *S* must decode Σ into *subject plus predicate*; then he must decode *subject* into *noun-phrase* and *prepositional phrase*; then he must decode *noun-phrase* into *article plus noun*; and only then is he ready to generate the first word, i. e., *The*. For that sentence the *S* would have to make seven decoding operations before the word *The* could be generated. In the Fig. 1 sentence, on the other hand, the *S* would decode Σ into *subject-plus-predicate*; then decode *predicate* into *article-plus modified-noun*; and then he would decode *article* into the word *The*. For that sentence, there are only five decoding operations necessary before the word *The* could be produced. Therefore, one might expect that the latency between the signal which sets the occasion for emitting the sentence, and the actual emission of the first word of the sentence, should be less for the Fig. 1 sentence than the Fig. 2. sentence.

Another very similar aspect of the model has been pointed out by Yngve (1960) who has proposed a model of sentence generation which is the same as the decoding-operation hypothesis. He points out that as *S*s are going through the decoding operations necessary to produce a word they are storing items in their short-term memory. As the number of items stored in the short-term memory increases, the difficulty of handling the sentence should also increase. He has suggested that if one labels all the left branches in a tree diagram with the digit one, and all the right branches with zero, the number of items in short-term memory at any point in a sequence can be determined by starting with the top node and tracing through the tree diagram to the point in question and, in the process, counting the number of ones which were encountered. That number should equal the number of items in short-term memory at that particular point. Yngve has called that number the depth of the sentence. Martin (1965) has demonstrated that the ease of learning a sentence is, indeed, a function of its depth. Therefore, if Ss do generate sentences as the decoding -operation hypothesis suggests, the latency between the onset of the stimulus and the occurrence of the first word in the sentence should be a function of the depth of the first word in the sentence.

To test that hypothesis, two sets of sentences were constructed which were highly similar, but differed in the depth of the first word. One set of sentences conformed to the type A structure presented in Fig. 7, and the other set conformed to the type B structure. As can be seen, the two sentences contained exactly the same words; and the type B structure is produced by simply taking the third and fourth words of the type A structure and interchanging them with the fifth and sixth words. For the type A sentence, there are six decoding operations occurring prior to the one which produced the word The, and the depth of the word The is three. For the type B sentence, there are eight decoding operations occuring before the word The is generated, and the depth of the word The is four. In addition, the mean depth of all eight words in the type

A sentence is 1.6, whereas the mean depth of all eight words in type B sentence is 2.0. Therefore, not only should Ss have more difficulty learning the type B sentence, but the latency between the onset of the stimulus and the occurrence of The should be greater than is the case for the type A sentence. The following experiment was an attempt to test those expectations.

Method

The sentences used in the experiment were: 1) The pupil who fell behind us is good; 2) The parent who spoke with me is old; 3) The person who ran under them is thin; 4) The poet who stood near you is rich; 5) The pupil behind you who fell is good; 6) The parent with me who spoke is old; 7) The person under them who ran is thin; 8) The poet near you who stood is rich. Four different sentences were used and they could occur in either of the Fig. 7 structures. The first four sentences listed above conform to the type A structure, and sentences five through eight are the same sentences, but conforming to the type B structure. The Ss learned the sentences as paired-associate responses to digits, and each S learned all four sentences. Two of them occurred as a type A sentence and the other two occurred as a type B sentence. In no case did corresponding sentences occur in the same list. In addition, to the four sentences, each S also learned The and The patient as responses. Therefore, each S learned six pairs. The responses to four of the stimuli were sentences; to one stimulus it was The patient; and finally to one it was just The.

The method used was the standard procedure outlined in Chapter II. The Ss learned the sentences to a criterion of two perfect trials, and they were presented at a 4-second: 4-second rate.

After the Ss learned the paired-associate list, they were given 20 overlearning trials during which a Lafayette Voice Reaction Timer and a Hunter Klockounter were used. The experimenter stated the stimulus digit into one microphone which activiated the timer, and when the S stated his response into another microphone, the timer was turned off. In that way the time between the stimulus digit and the first word of Ss response was measured. The stimuli were presented at a rate of approximately one every eight seconds. During the overlearning trials, then, a record was kept of the latency between the stimulus and the first word of the sentence the Ss emitted.

The word The is not sufficiently powerful to reliably trip the relay in the timer. Therefore, the first noun in the sentence always started with the letter P. If the Ss gave an incorrect response, or his response was stated too softly to trip the relay, his time was not recorded and two or three items later the stimulus was presented again.

The Ss were 24 male introductory psychology students, and they were assigned to one of the six groups in alternating order as they appeared for the experiment. During the original learning task, a verbatim record was kept of the Ss' responses, but during the overlearning trials the only data recorded were the latencies.

Results and Discussion

Transitional error probabilities (TEPs) in original learning. The TEPs for each transition for each type of sentence are given in Table 9 along with the number of hypothetical decoding operations the model predicts at each transition. In that the Ss learned the list to a criterion of two perfect trials, the number of trials upon which there was available data varied from S to S .

The S that learned the list the most rapidly did so within 12 trials so the TEP data reported in Table 9 are based on only the first 12 trials.

TABLE 9

	Transition Number						
	1	2	3	4	5	6	7
Type A Sentence							
Number of Operations	1	10	7	4	1	4	1
TEP	.0825	.1740	.1760	.1740	.1740	.0542	.0810
Type B Sentence							
Number of Operations	1	4	1	4	1	4	1
TEP	.1580	.7000	.3250	.5000	.2320	.4160	.0950

A Friedman two-way analysis of variance was used to determine if the seven TEPs for each sentence type were reliably different. For the type A sentences, $X^2_r = 15.85$ (df = 6, $p < .02$), and for the type B sentences, $X^2_r = 35.08$ (df = 6, $p < .001$). The variation from transition to transition in the magnitude of the TEPs did seem to be reliable.

To determine the extent to which the relative magnitude of the TEPs approximated that predicted by the decoding-operation hypothesis, the correlation (rho) was computed between the rank order of the TEPs and the rank order of the decoding operations for the transitions in each of the sentence types. For the type A sentences, rho = .50 and for the type B sentences, rho = .87. As in the previous studies, the significance of the correlations was determined by computing the rho between the obtained and predicted TEPs for each sentence type and for each S. For the type A sentences there were 21 positive and 3 negative rhos ($p < .001$), and for the type B sentences there were 22 positive and 2 negative rhos ($p < .001$).

One interesting characteristic of the materials used in the present experiment is that while the words in the two sentence types are identical, there is a rather subtle structural distinction between them. That distinction allows for the possibility of determining the extent to which the TEP pattern is sensitive to such structural distinctions. To do this a correlation was computed for each Ss between the TEPs he obtained for each sentence type and both the pattern predicted by the type A structure and the pattern predicted by the type B structure. If the TEPs are sensitive to slight structural distinctions, then the correlation between the pattern and that predicted from the correct structure should be significantly greater than the correlation with that predicted from the incorrect structure. For the type A sentences the TEP pattern correlated .50 with that predicted from the correct structure and -.15 with that predicted from the wrong structure. For the type B sentences, the TEP pattern correlated .87 with the correct structure and .79 with the pattern predicted by the incorrect structure. For both sentence types, 17 Ss had a higher correlation between the TEP pattern and the correct structure than for the incorrect structure ($p = .032$).

Rate of original learning. Yngve (1960) has suggested that sentences which differ in depth should differ in the ease with which Ss can handle them. To test whether the two sentence types were learned at different rates, each S was scored for the mean trial on which each sentence type occurred correctly for the first time during original learning. The type A mean was 10.77, and for type B it was 15.83 ($t = 6.49$, $df - 23$, $p < .001$). The deeper type B sentences were clearly more difficult to learn. An examination of the TEPs indicates that those for the type B sentences exceed those for the type A sentences on every transition. In addition, the discrepancies seem to be sizeable. That suggests that the significant difference in rate of learning for the two sentence types was not only the result of differential ease of pairing the sentence responses with their appropriate stimuli, but they also differed in the ease with which the sentence responses were integrated.

Overlearning trials. The latency between the stimulus and the first sound in the response, as a function of the number of overlearning trials, is given in Table 10 for both sentence types. For both sentence types, the drop in latency from trial 1 to trial 20 was significant (type A, $F = 5.20$, $df = 9/207$, $p < .001$; type B, $F = 5.36$, $df = 9/207$, $p < .001$). The mean latency for the type A sentences during overlearning was 1.157 seconds as opposed to 1.209 seconds for the type B sentences ($t = 1.77$, $df = 23$, $p < .05$). As predicted, the latency for the type B sentences was greater. The experiment was replicated, and in the replication the obtained means were 1.180 and 1.264 ($t = 2.47$, $df = 23$, $p < .01$).

TABLE 10

Sentence			Trials							
Types	1–2	3–4	5–6	7–8	9–10	11–12	13–14	15–16	17–18	19–20	All Trials
Type A	1.401	1.206	1.215	1.162	1.184	1.094	1.082	1.078	1.089	1.061	1.157
Type B	1.465	1.385	1.235	1.206	1.145	1.156	1.236	1.097	1.103	1.067	1.209
Both Groups	1.433	1.295	1.225	1.184	1.165	1.125	1.159	1.087	1.096	1.064	1.183

One possible explanation for the difference in latency between the two sentence types could be the fact that they were learned at unequal rates during original learning. While the learning criterion was two perfect trials, that may have allowed several overlearning trials for the type A sentences. If that were the source of the latency difference, then there should have been a fairly sizeable difference between the latency for the type A and B sentences early in the overlearning trials, with the difference decreasing as they progressed. That is, during the overlearning trials there should have been a significant trials by sentence-type interaction for the latency data. However, the interaction was not significant ($F = 1.02$, df - 9/207, $p > .25$). Furthermore, the mean latency for the last 10 overlearning trials (.051 seconds) is about the same as for the first 10 trials (.054 seconds). It would be difficult to reconcile that result with an explanation of the significant mean difference that is based on differences in degree of original learning. Those results are also supported by the replication data.

The latency between the onset of the stimulus and the occurrence of the first word in the response should have been occupied by the time to 1) execute the association; 2) perform the initial operations to generate the response; and 3) generate the first response. *The patient* and the two sentence types differ only in terms of the number of initial operations, with *The patient* having zero operations; the type A sentences having six; and the type B having eight operations. If the latency between the signal and the first word of the response is a linear function of the number of decoding operations necessary to produce it, then if the mean difference in latency between the type A sentences and the type B sentences is multiplied by three and added to the mean latency for *The patient*, the result should be a reasonable approximation to the latency for the type A sentences. That is, if the difference between the type A sentences and the type B sentences is the time to perform two operations, then three times that difference should be the time to perform six operations. The predicted value of 1.091 seconds differs from the obtained value of 1.157 (± .057) by .066 seconds. In light of the standard error of the obtained latency, the fit is fairly good.

The results of the experiment again seem to support the decoding-operation hypothesis, and it does so in several ways. First, the TEP data for each sentence type conform to that predicted by the decoding-operation hypothesis. Second, the results indicate that the TEP pattern is sensitive to relatively small structural changes. Third, the original learning data conform to Yngve's (1960) depth hypothesis to the extent that the deeper sentence type was the most difficult to learn. Finally, the latency data for the onset of the sentences was related to the number of hypothetical operations at that point.

Experiment XI

If the decoding-operation hypothesis and the unit hypothesis are combined and it is predicted that the TEPs are a function of the number of decoding operations at each transition, then it is necessary to assume that Ss are decoding ahead of their verbal production. That is, the assumption that Ss do not begin to generate a unit they cannot complete, implies the further assumption that Ss make the decoding decisions necessary to produce the unit prior to its actual production. Certainly, such decisions would be necessary for the S to know that he cannot complete the unit. If that is the case, Ss should make the decoding decisions necessary to produce a unit at the unit boundary immediately prior to the unit. One important implication of that assumption is that if difficulty should be introduced somewhere within a unit, or the length of a unit is increased, the behavioral effect of such manipulations should not occur at the point at which they are introduced, but, rather, at the transition from the last element of the previous unit to the first element of the unit in question.

Experiments 9 and 10 were an attempt to determine whether the processes Ss use to generate a sentence conform to that hypothesized by the decoding operation model. Experiment 11 represented an attempt to determine the location within the sequence that the hypothetical processes, decisions, or decoding takes place. The general procedure was to have Ss learn letter sequences as paired-associate responses, similar to those used in experiment 5. The data from experiment 5 indicate that if blank spaces are left between the letters on presentation, Ss will use these blank spaces to break the sequences into subunits. The effect is evident in the TEP spikes on the transitions where a blank was introduced. The present experiment was an attempt to determine whether the magnitude of the TEPs on one of these between-unit transitions would be a function of the number of elements or letters within the unit following the transition. That is, if, as assumed, Ss make all of the decoding decisions necessary to produce a unit before attempting to produce any of it, and they do not produce units they cannot complete, then the TEP on the transition to that unit should be an increasing function of the size of the unit. In addition, although the data from experiment 5 indicate that the within-unit TEPs increase as the size of the unit increases, the unit hypothesis clearly implies that the rate of such increase should be greater on the transition to the unit than on any transition within the unit. Finally, if the units within a sequence are relatively independent, then increasing the size of one subunit should have little or no effect on TEPs within other subunits. For example, if one group of Ss learns a sequence like NG V as a paired-associate response, another group learns NG VH as a response, a third group learns a sequence like NG VHS, and a fourth group learns NG VHSB, it would be expected that from Group 1 to Group 4 there would be a marked increase in TEP on the G-V transition; but little or no change on the N-G transition or the V-H transition.

Method

Materials and design. The Ss' task was to learn three paired-associates which had digits as stimuli and letter sequences as responses. The letter sequences were the following: 1) NG VHSB; 2) TC KFZW; 3) LY QJMX; 4) SB JHFZ; 5) NY QDRC; 6) MK XVGW. The sequences were constructed using the same method as for experiment 5.

The Ss were divided into four major groups which differed in terms of the number of letters in the sequences they learned. All four groups learned sequences which consisted of two letters followed by a space. They differed, however, in the number of letters which followed the space. Group 1 had, as responses, only the first three letters of each sequence; Group 2 had the first four letters of each sequence as responses; Group 3 had the first five letters as responses; and Group 4 had all six letters as responses. Half of the Ss in each of the four groups learned the letters from the first three sequences listed above, and the other half of the Ss learned the last three sequences. For all groups, when the letter sequences were presented to them during learning, a blank space was left between the second and third letters.

The standard procedure described in Chapter II was used for the experiment. The materials were presented at a 4-second: 4-second rate, and all Ss received a constant 20 learning trials. As in the previous studies, their responses were recorded verbatim, and the TEPs were computed. The Ss were 64 introductory psychology students, and they were assigned to groups and lists in alternating order as they appeared for the experiment.

Results and Discussion

The TEPs for each transition for each group are given in Table 11, along with the unconditional probability that the first letter of the sequence was wrong (i. e., transition zero). That unconditional probability is the TEP for the transition from the stimulus digit to the first letter of the sequence.

TABLE 11

Groups	Transitions					
	0	1	2	3	4	5
I	.12	.03	.02			
II	.26	.06	.12	.08		
III	.21	.06	.24	.12	.07	
IV	.19	.02	.35	.14	.22	.09

The difference between the two intrasequence TEPs for Group I was tested using a Wilcoxon matched-pairs signed-ranks test, and the difference was not significant ($T = 19.5$, $N = 10$, $p < .05$). For the other three groups, the hypothesis of intrasequence TEP equality was tested using a Friedman two-way analysis of variance and for all three groups the hypothesis of equality was rejected (Group 2: $X_r^2 = 6.28$, df = 2, $p < .05$; Group 3: $X_r^2 = 15.00$, df = 5, $p < .01$; Group 4: $X_r^2 = 6.00$, df = 4, $p < .001$).

To determine whether the Ss were using the blank space as a unit marker, the TEP on transition two was compared witn the mean of the other transitions for each group. If the blank space was used as a unit marker, the TEP on transition two should be greater than for the other transitions. Each S was scored for his TEP on transition two, and the mean of the other transitions, and the difference was tested using the Wilcoxon. For Groups I and II the difference was not significant; but it was for Groups III and IV (Group 3: $T = 1$, $n = 16$, $p < .01$; Group 4: $T = 0$, $n = 16$, $p < .01$). A possible explanation for the failure to obtain a significant difference for Groups 1 and 2 could be that the sequences were so short that the Ss were able to learn them as a single unit with very little difficulty.

The TEPs for transitions 0, 1, 2, and 3 are plotted in Fig 8 as a function of the number of letters in the second subunit of the sequence. The influence of the size of the second subunit on the TEP at each of the transitions was tested using a Kruskal-Wallis one-way analysis of variance. For transition 0, $H = 39.7$ ($df = 3$, $p < .001$); for transition 1, $H = 8.4$ ($df = 3$, $p < .05$); for transition 2, $H = 40.1$ ($df = 3$, $p < .001$); and for transition 3, $H = 2.6$ ($df = 2$, $p > .05$). The difference in TEP, then, was significant for transitions 0, 1, and 2, but not for transition 3. For all of the curves, the differences between succeeding points were tested using a Mann-Whitney U-test, and all of the differences between succeeding points were significant for transition 2. The only other significant differences were between Groups 1 and 2 on transition zero and between Groups 3 and 4 on transition one.

To test whether the slopes of the curves were significantly different from one another, each S was assigned three scores; 1) the signed difference in TEP between transitions 1 and 2; 2) the signed difference in TEP between transitions 1 and 3; and 3) the signed difference in TEP between transitions 2 and 3. If the slopes of the curves are significantly different from one another, then there should be a significant increase in the difference scores between the curves from Group 1 to Group 4. The reliability of the increase in mean difference in each case was tested using the Kruskal-Wallis and for the difference between transitions 1 and 2, $H = 38.6$ ($df = 3$, $p < .001$), and for the difference between 2 and 3, $H = 12.4$ ($df = 2$, $p < .001$). In addition, the increasing difference between transitions 1 and 3 was significant beyond the .05 level ($H = 6.3$, $df = 2$).

The one unexpected finding in the experiment was the significant H for transition 1. It had been anticipated that there would be no significant change in TEP on that transition correlated with the size of the second subunit. A possible explanation for the result can be made if it is assumed that: 1) there was a tendency for Groups 1 and 2 to treat their sequences as a single unit, as the data suggest, and 2) there is a slight increase in the magnitude of within-unit TEPs as the size of the unit increases, as suggested by the results of experiment 5. Both assumptions do have empirical support. If Groups 1 and 2 did treat their sequences as single units, then there should have been an increase in the TEP on transition 1 because the number of letters within the sequence increased from three to four from Groups 1 to 2. However, if Groups 3 and 4 used the space as a unit break, as the data suggest, there should have been a decrease in size of the unit in which the first transition appeared from Group 2 to Groups 3 and 4. Consequently, there should have been a decrease in the TEP for transition 1.

One further empirical check on that explanation can be made by examining the function for transition 0. If Groups 1 and 2 handled their sequences as consisting of a single unit, but Groups 3 and 4 used the space between the second and the third letters as a unit marker, then the size of the first unit for Group 1 was three letters; for Group 2 it was four letters; for Group 3 it was two letters; and for Group 4 it was also two letters. If, as the data of experiment 11 seem to indicate, the TEP on the transition to a unit is a function of the number of letters or size of the unit, then the TEP on the transition to the first unit (i. e., transition 0) should go up from Group 1 to Group 2 followed by a drop from Group 2 to Group 3, with Groups 3 and 4 approximately equal. An examination of Fig. 8 reveals that just such a function for transition 0 was in fact obtained. The apparent Transition by Group interaction for transitions 0 and 2 was tested by assigning each S, as a score, the signed difference in TEP between these two transitions. The results were analyzed using the Kruskal-Wallis and the apparent crossover of the two curves was reliable ($H = 17.44$, $df = 3$, $p < .001$). In addition, the predicted increase in TEP on transition 0 from Group 1 to Group 2 was tested using the Mann-Whitney U-test, and, while it was significant ($U = 62$, $N_1 = 16$, $N_2 = 16$, $p < .01$), the drop in TEP for that transition from Group 2 to 4 was not significant ($H = .52$, $df = 2$, $p > .05$).

The failure to get a significant drop in TEP from Group 2 to Group 4 could indicate that when the stimulus digit appears the Ss not only decide what elements they are going to produce for the first subunit, but they also make some decisions regarding the units following the first subunit. If it is assumed that the Ss generate the letter sequences in a manner similar to that which the decoding-operation hypothesis supposes, then it might be expected that when the stimulus digit appears, the Ss first decode the sequence into its subunits, storing all but the first subunit in short-term memory, and then decode the first subunit into the letters it contains. If that were the case, then the failure to get a marked drop in TEP on transition 0 from Group 2 to Groups 3 and 4 could be explained.

In general, the results of experiment 11 support the unit hypothesis and the decoding-operation hypothesis to the extent that the major behavioral effect of increasing the size of a unit or a subunit does not occur within the particular subunit, nor within other subunits within the same sequence. It does occur, however, on the transition to the unit in question. At the very minimum, it would appear that an explanation for the data of experiment 11 must presuppose that the Ss are making response decisions prior to their actual emission.

Experiment XII

The preceding few experiments were designed to investigate some of the more critical implications of the decoding-operation hypothesis. Of the implications, the most critical ones, as far as a description of language production is concerned, seem to be: 1) those which describe the nature of the associative system Ss use in generating language sequences; 2) the fact that if Ss do proceed through the system of decoding operations, there should be time or latency factors correlated witn the number of decoding decisions; and 3) the location and inclusiveness of the decoding decisions an S makes at any point within a sequence. Experiment 9 and the Frankart (1964) study yielded data which were in accord with the associative system the decoding-operation hypothesis assumes Ss use when producing a sentence. Experiment 10 and the data reported by Maclay and Osgood (1959) do indicate that latencies are related to the number of decoding operations the model supposes should occur at various points in the generation of a language sequence. Finally, experiment 11 indicates that the response decisions that Ss make in generating a sequence conform to those predicted by the hypothesis, both in terms of their location and the inclusiveness of the decision. These data, then, offer rather general support for the use of the hypothesis as a general model of language generation or production. While it is certainly the case that the hypothesis has implications beyond those subjected to experimental test, those implications which have been supported appear to be the model's most critical elements.

If the above implications of the hypothesis represent an explanation for the TEP pattern, as the data suggest, an interesting question which arises concerns when Ss learn to handle language sequences in such a manner. That is, the underlying structure of a sequence should define the units to which the above implications apply and, consequently, should define the resulting TEP pattern. Therefore, it would be interesting to examine children's developing knowledge of grammar or structure by looking at the degree to which their TEP patterns approximate the ideal pattern predicted from the structure (or the pattern obtained by adults when learning the same material). Most measures of an S's knowledge of grammar are dependent upon his ability to verbalize rules. However, the fact that children's verbal behavior can conform to grammatical rules prior to their ability to verbalize the rules, indicates they have a functional knowledge of the rules prior to their ability to verbalize them. The correlation between an S's TEP pattern and that predicted from the structure might be one way of evaluating his functional knowledge of grammar, without requiring him to verbalize rules. Experiment 13 was a master's thesis by Edward Mathews (1965), and it attempted to examine the development of TEP patterns during early school years.

Most of the prior experiments which attempted to evaluate children's functional knowledge of grammar have dealt exclusively with their knowledge of parts-of-speech. That research, however, has yielded an interesting relationship between age and functional knowledge. These experiments have used both children's associations and various usage tests. One interesting result of these experiments is a demonstration that children change from giving syntagmatic associations to paradigmatic associations during the early school years. That is, if young children are given a stimulus word like deep, they show a marked tendency to give response words from a different word class (e. g., hole). Older children and adults, however, show a greater tendency to give a response word which comes from the same part-of-speech as the stimulus (e. g., shallow). The shift from syntagmatic to paradigmatic associations has been attributed to the children's increasing functional knowledge of grammar. Most of the data has indicated a relatively rapid shift from first grade through fourth grade, with fourth graders showing a set of responses similar to those obtained on adults. At least, the change from fourth grade to adults is nowhere near as striking as is the change from first grade to fourth grade.

Another technique which has been used is the usage test. Basically, that technique involves introducing nonsense words in sentence positions which define them as belonging to various parts-of-speech. The child is then asked to use the new word in a sentenceof his own. The question asked by the technique is the extent to which the children can use the sentence context to define the part-of-speech to which the nonsense word belongs, and then demonstrate that knowledge by using the nonsense word appropriately in a new sentence. The Ss are simply scored for whether they use the nonsense word in the same manner as the experimenter. The results of various experiments using this technique demonstrate a relationship between age and correct usage which is almost identical to the function which relates percentage of paradigmatic associations and age. By combining the data from these two sources (. i. e., children's word associations and usage tests), it seems clear that there is a rather rapid acceleration in children's functional knowledge of grammar during the early school years (i. e., from grades one through four). Experiment 12 attempted to examine the generality of that conclusion by determining whether the function obtained for parts-of-speech is also obtained for functional knowledge of phrase-structure rules, as inferred from the TEP pattern. It was hypothesized that there should be a fairly rapid increase from grades one to four in correlation between an S's TEP pattern and both that predicted from the structure and that obtained on adults. The pattern obtained with fourth graders should look very similar to that obtained on adults.

Method

Again the standard paired-associate procedure described in Chapter II was used, but there were a few important modifications. The Ss learned nine digit-sentence pairs. The sentences conformed to the structure illustrated in Fig. 2 (e. g., The house across the street burned down), and the words used in the sentences were taken from the first grade reader used in the school system from which the children were drawn. The materials were presented to the Ss on cards which the experimenter turned by hand, rather than using the Lafayette Memory Drum as in the previous experiments. On a single card, a stimulus digit was printed on the left half of the card; and on the right half of the card the response sentence was typed. During stimulus presentation, the right half of the card was covered with a 3" x 5" card which was taped at the top to form a hinge. The stimulus was presented alone for approximately 4 seconds, and then the experimenter turned up the 3" x 5" card exposing the sentence response for another 4 seconds. Then the entire card was turned over and a new stimulus digit appeared for 4 seconds, followed by a 4-second response exposure, etc.

The Ss received 14 learning trials with presentation at approximately a 4-second: 4-second rate.

The procedure used for the first graders was similar to that used for the second, third, and fourth graders except that the stimulus digits wereprinted on 5" x 8" cards and the responses were presented orally.

The Ss were 80 children, 20 each from the first grade, second grade, third grade, and fourth grade. The school from which the sample was drawn was in a middle class neighborhood in a small town in central Ohio.

Results

Rate of learning. The mean number of correct words emitted per sentence during the course of learning for each of the four grades were: Grade 1, 1.82; Grade 2, 2.41; Grade 3, 2.49; Grade 4, 3.16. The significance of the increase in mean number of correct words per sentence across the four grades was tested using a Kruskal-Wallis one-way analysis of variance, and it was significant beyond the .05 level (H = 8.47, df = 3). The significance of the difference between individual grades was tested using a Mann-Whitney U-test, and the difference between Grades 1 and 2 and the difference between Grades 3 and 4 were significant beyond the .05 level. The difference between Grades 2 and 3 was not significant. Ease of learning did appear to be positively related to age.

A second measure of rate of learning was obtained by computing the mean TEP per grade. The respective means were .27, .17, 10, and .07. The decrease in TEP from Grade 1 to 4 was significant beyond the .001 level (H = 25.31, df = 3). The difference between the grades were tested using a Mann-Whitney U-test, and the difference between Grades 1 and 2 and the difference between Grades 2 and 3 were both significant beyond the .05 level. The difference between Grades 3 and 4, however, was not significant. The significant drop in TEP from Grade 1 through Grade 4 indicates that the increasing ease with which the Ss handled the learning task, as age increased, must be attributed, at least partly, to increasing ease in integrating the sentences.

Transitional error probability patterns. The TEPs for each of the six transitions, for each grade, are given in Table 12. For each grade, the hypothesis that the six TEPs were equal was tested using a Friedman. For each grade, the differences were highly significant (Grade 1, $X_r^2 = 59.35$, df = 5, $; < .001$; Grade 2, $X_r^2 = 67.86$, df = 5, $p < .001$; Grade 3, $X_r^2 = 52.39$, df = 5, $p < .001$; Grade 4, $X_r^2 = 56.62$, df = 5, $p < .001$).

For each S, and for the mean of each grade, a correlation was computed between the obtained TEP pattern and both that predicted by the decoding-operation hypothesis and the pattern obtained in experiment 3 using adults. The correlation for each grade between the obtained TEP pattern and that predicted from the structure was: Grade 1 = .61; Grade 2 = .61; Grade 3 = .92; and Grade 4 = .92. All four correlations were significant beyond the .01 level, with every S showing a positive relationship between the obtained and predicted TEP pattern. In addition, the four correlations were significantly different from one another, and the increase from Grade 2 to Grade 3 was significant beyond the .05 level ($U = 133.5$, $N_1 = 20$, $N_2 = 20$).

The correlations for the four grades between the obtained TEP patterns and that obtained with college sophomores were: Grade 1 = .71; Grade 2 = .77;

TABLE 12

Grade	Transitions					
	1	2	3	4	5	6
I	.3785	.4331	.3129	.0258	.1942	.0140
II	.2800	.3027	.1003	.0202	.1245	.0073
III	.0844	.2390	.0922	.0235	.0995	.0086
IV	.0599	.1736	.0774	.0235	.0810	.0036

Grade 3 = .91; and Grade 4 = .94. Again, the four correlations were significantly different from one another, and the increase from Grade 1 to Grade 2 and the increase from Grade 2 to Grade 3 were both significant beyond the .05 level.

Discussion

The current experiment was based on the assumption that a TEP pattern can be used as an estimate of an S's functional knowledge of phrase-structure rules. At the very least, it is reasonable to suppose that the TEP pattern does reflect an Ss ability to use these rules during learning, and that should be related to his functional knowledge of the rules. The question examined by the present experiment concerns the form of the function relating age and the degree to which an S's TEP pattern approximates the ideal predicted by the structure.

The prior research on the acquisition of speech categories (i. e., parts-of-speech) indicates a rather rapid increase in children's functional knowledge of these categories from age 6 (first grade) to age 9 (fourth grade). By age 9, the children's performance on the tests approximated the performance of adult Ss. If those results reflect an increasing functional knowledge of grammar, then the same function should be obtained for the relationship between age and TEP pattern.

The results of experiment 12 were in accord with the expectation. The negatively-accelerated function relating age of children and the correlation between the TEP data for the children and for adults, is very similar to the negatively-accelerated function between age and performance on the word usage tests in the previous studies.

A small experiment was also conducted to determine whether there would be a relationship between grades in English and correlations between expected and obtained TEP patterns. The Ss were 35 students in a tenth grade English class from a parochial school. The experimenter read the numbers and their associated sentences to the Ss and then read each number and asked the Ss to write down as much of the appropriate sentence as they could recall. The procedure continued for 14 study-and-test trials. The TEPs were computed and the students were rank ordered for both grades in English and correlation between obtained and expected TEP pattern. The two sets of scores were correlated and the result was only .09.

Experiment XIII

The specific purpose of experiment 13 was to examine the various kinds of errors Ss make as they are producing the sentences during learning. The unit hypothesis predicts a recall dependency between adjadent elements from the same subunit, and a somewhat lower recall dependency between elements from different subunits. Specifically, the hypothesis predicts that if two adjacent elements are from the same subunit, then if one occurs, the other should also occur; if one element does not occur, the other should not occur either. The unit hypothesis, then, predicts a bidirectional dependency spanning transitions within the same subunit. If the element before the transition occurs, the probability of an error on the element after the transition should be about the same as the probability of a correct response on the element after the transition, given an incorrect response on the element before the transition.

The elaboration of the unit hypothesis made by the decoding-operation model results in a unidirectional dependency prediction. That is, the model predicts that given the S̲ has emitted the item before a transition correctly, the probability that he will correctly emit the item after the transition should be a function of the size of the unit that begins with the word after the transition. On the other hand, if an S̲ has emitted the second element of a pair correctly, the probability that the element before the transition was also correct should be a function of how readily the S̲s can form an association between the two words. If, as has been assumed, the ease with which these associations can be formed is approximately constant throughout the sentence, these reverse TEPs should be unrelated to the pattern of errors predicted by the decoding-operation model. Therefore, the unit hypothesis predicts that the pattern of bidirectional dependencies should conform to the pattern of decoding operations. The decoding-operation model, on the other hand, predicts that the probability of an error on the word after a transition, given the word before the transition was right (P E/C), should conform to the pattern predicted from the model, but the probability of a correct on the word after a transition, given an error on the word before the transition P(C/E), should not necessarily be related to the predicted pattern. Experiment 13, then, was designed to test the hypothesis that the pattern of hypothetical decoding operations should be more closely related to the pattern of TEPs, or P (E/C), than to the reverse TEPs or P (C/E).

Method

The standard paired-associate procedure described in Chapter II was used for the experiment. Digits were used as stimuli, and eight sentences conforming to the Fig. 2 structure were used as responses. The sentences were: 1) The man near the car looked sick; 2) The street by your house needs fixing; 3) The girls who live here work hard; 4) The cat under his window ran away; 5) The people that went swimming left early; 6) The box from the store arrived today; 7) The dogs with black tails sat quietly; 8) The child in her yard picked flowers. The materials were presented to the Ss at a 4-second: 4-second rate, with a 4-second intertrial interval. All Ss received 13 learning trials.

The Ss were 20 introductory psychology students. Their responses were recorded verbatim.

Results

The results of the experiment are presented in Table 13. As in the previous

TABLE 13

Type of Error	Transition Number					
	1	2	3	4	5	6
P E/C	.057	.214	.082	.022	.109	.025
P C/E	.010	.067	.098	.095	.088	.032
1 - phi	.08	.29	.20	.15	.21	.03

studies using the Fig. 2 structure, the correlation between the number of hypothetical decoding operations at each transition and the TEP (P E/C) was .92 ($p < .05$). The correlation between the number of decoding operations at each transition and the reverse TEPs (P C/E) was .39 with 18 of the 20 Ss having positive correlations ($p < .05$). To determine whether the two correlations were significantly different from one another, a comparison was made, for each S, between his two correlation coefficients. For every S, the correlation between the number of hypothetical decoding operations at each transition and the TEP was greater than for the reverse TEPs. In addition, the correlation between the forward and the reverse TEPs was only .03. Although the correlation of .39 between the predicted TEP pattern and the reverse TEPs was somewhat larger than expected, the fact that it was significantly lower than the value obtained using the forward TEPs, and the fact that the pattern of forward and reverse TEPs correlate only .03, demonstrates that the two sources of independence are independent of one another and that the forward TEPs approximate those predicted by the model somewhat more closely than do the reverse TEPs.

One way to take both sources of independence into consideration at the same time would be to compute a phi coefficient for each transition. The phi coefficient would yield a measure of dependence between the two elements surrounding a transition and, consequently, (1 - phi) would be a measure of independence similar to that obtained with the TEPs, except that it takes both sources of independence into consideration. As can be seen by examining the bottom row ot Table 13, the values of (1 - phi) for each of the six transitions do approximate that predicted by the decoding operation model. However, in that the forward TEPs correlate significantly better with the model than do the reverse TEPs, it might be reasonably argued that the correlation between the values of (1 - phi) and the decoding operation model can be accounted for by the forward TEPs.

Experiment XIV

The previous sequence of experiments was initiated in an attempt to understand the way in which Ss use structure when learning language materials. One of the first studies demonstrating the phenomenon was that reported by Miller and Selfridge (1950). Their data indicate that Ss' rate of learning language sequences is proportional to the degree to which the sequences approximate English. Their interpretation of the result was that the higher-order approximations to English allowed the Ss to use previously-established short-range associations. As indicated in the introduction to the present report, research since that time has suggested the explanation may not be that simple.

Both Yngve (1960), and the experiments reported above, suggest that Ss may use tree diagrams (P-markers) when attempting to either recall or generate a sentence. The data reported above seem to indicate that the difficulty Ss have in integrating adjacent word pairs within a sentence is related to the structural constraints implied by the sentence's P-marker.

Swanson (cited by Osgood, 1957) and Epstein (1961) found that if a sequence of word-like nonsense forms is inflected so that a P-marker can be assigned to it (e. g., The maff velms oothly um glox nerfs) it will be learned more readily than if it is not inflected (e. g., The maff velm ooth um glox nert). Rosenberg (1964), however, has questioned these results for, among other reasons, the fact that they used sequences in which there was intrasequence formal similarity and intraposition similarity across the sequences. In addition, they used rather long sequences. Rosenberg's Ss learned short sequences which had low association value CVC trigrams as the word roots, and within a sequence no letter was used more than once. He also used naive Ss. His data indicate that putting the CVCs in sentence frames (e. g., The vun zots beped mivly) does not increase the ease of learning the sequences.

One way of reconciling Rosenberg's results with the earlier data is to suggest that the step of assigning a P-marker to a sequence tends to insulate it from intrasequence interference and interference from prior language experience. When Rosenberg reduced the possibility of interference effects the facilitation from structure seemed to disappear. Also, the clearest and most consistent evidence of facilitation from structure is obtained in experiments which use grammatical vs. random sequences of English words. Interference should be greater when English speaking Ss are learning random sequences of English words than when naive Ss are learning sequences of CVCs.

Keppel and Underwood (1962) have demonstrated that the short-term memory (STM) of CCCs is influenced by accumulating interference. Therefore, if Ss are presented a series of CVC sequences which they are asked to recall, their ability to recall them should be a decreasing function of the number of previous sequences they have recalled. The above considerations would suggest, however, that if Ss can assign a P-marker to a sequence of CVCs the influence of the accumulating interference should be reduced.

Method

Materials and procedure. The Ss were presented a 3" x 5" card upon which was typed either a string of five CVCs (e. g., jup teb vec caz ged) or five CVCs imbedded in a sentence frame (e. g., The jupish tebs had veced the cazly geders). The Ss were instructed to read aloud the material on the card and then the card was removed. The S then counted backward from five to one and immediately attempted to recall the material on the card. His response was recorded verbatim, and a new card was then presented. Each S waspresented 32 such sequences: 16 were just strings of 5 CVCs (strings) and 16 were 5 CVCs imbedded in a sentence frame (sentences).

The 32 sequences of 5 CVCs were constructed using the Archer (1960) tables. All the CVCs were pronounceable and had association values between 30% and

60%. Four different sentence frames were used: The ______-ful ______ near ______ ______ is ______-ing; The ______ ______ ______-ed under the ______ ______-ly; The ______-ous ______-s are ______-ing in the ______-ed ______-s; and The ______-ish ______-s had ______-ed the ______-ly ______-ers. Two of the frames have only two of the five CVCs inflected and two have all five CVCs inflected. Four CVC sequences were randomly assigned to each sentence frame, with the restriction that the final consonant of a CVC could not be inconsistent with its inflection (e. g., a CVC ending in s could not be used if the inflection was a plural -s).

Design. The order of presentation of the 32 items was determined by organizing the list into 4 blocks of 8 items each. Within each block of eight there were four CVC strings, plus one sentence representing each of the four frames. The specific strings and sentences that appeared in each block of eight, and the order of presentation within a block, was determined randomly. The items were presented to the Ss continuously, with no distinction made between blocks.

A second group of Ss received the same CVC sequences and frames with the exception that the sequences that were presented alone for the first group (strings) were imbedded in the frames (sentences) for the second group, and vice-versa. In addition, each of the two major groups had two subgroups. For one subgroup the order of strings and sentences within a block was determined randomly. For the other subgroup the positions of the sentences was interchanged with the positions of the strings. That step was taken to insure that strings and sentences would occur equally often in each of the 32 positions within the list.

Scoring. Each S̲s was given four scores for each of the four blocks of eight (i. e., 16 scores in all). The first score was the mean number of CVCs correctly recalled from the strings; the second was the mean number of CVCs correctly recalled from the sentences with five inflections; the third was the mean number of CVCs recalled from the sentences with two inflections; and the fourth was the mean number of CVCs recalled from all the sentences.

Subjects. The S̲s were 40 introductory psychology students who participated as a part of their course requirement. An equal number of S̲s were assigned to each of the four subgroups in alternating order as they appeared for the experiment.

Results

The results of the experiment are presented in Table 14. The values within each cell are the mean number of CVCs correct for each string, sentence, or sequence within each block of eight. As can be seen in the bottom row of the table, the mean recall scores decreased from the first block to the last ($F = 3.60$, $df = 3/117$, $p < .02$).

The mean number of CVCs per sequence recalled for the sentences was 1.79 and for the strings the mean was 1.88 ($F = 3.14$, $df = 1/39$, $p > .05$). Not only did the S̲s not recall more CVCs from the sentences, but the difference that did occur favored the strings. For the materials and procedures used, there did not seem to be a facilitating effect on recall from the sentence frames, which supports the data reported by Rosenberg (1964).

TABLE 14

Sequence Type	Blocks				All Blocks
	1	2	3	4	
Sentences					
5-inflected	1.93	1.88	1.95	1.76	1.87
2-inflected	2.00	1.46	1.51	1.84	1.70
All sentences	1.96	1.66	1.74	1.79	1.79
Strings	2.06	2.03	1.83	1.61	1.88
All sequences	2.01	1.85	1.78	1.70	1.83

The hypothesis outlined above predicts that there should be an accumulating interference effect demonstrated, and although the prior data indicates there may not be an overall difference in recall of the CVCs from the strings and sentences, there should be a difference in the effect of the interference on the two kinds of sequence. An examination of the third and fourth rows of Table 14 suggests such a differential effect did occur. The interaction between blocks and sequence type (i. e., strings vs. all sentences) was significant beyond the .01 level (F = 4.0, df - 3/117). As predicted, the decrease in recall scores for the CVC strings was significant (F = 31.12, df - 3/117, $p < .001$) whereas there was no such significant decrease in recall scores for the CVC sentences (F = 2.24, df = 3/117, $p > .05$). One effect of the sentence frames, therefore, seemed to be that of reducing the effect of accumulating interference.

Two different types of sentence frames were used in the experiment. They differed, primarily, in the number of inflected CVCs. The mean recall scores for the two types of frames are given in the first and second rows of Table 14. The mean number of CVCs recalled from the sentences with five inflections was 1.87 while it was only 1.70 for the sentences with two inflections (F = 10.45, df - 1/117, p = .005). Although the main effect of blocks was not significant, there was a significant interaction between blocks and sentence type (F = 13.31, df = 3/117, p = .001). By examining the first two rows in Table 14 it can be seen there was little or no change across the four blocks in the Ss' ability to recall the CVCs from the sentences with five inflections (F = .47, df = 3/117), but there was a marked variation in their ability to recall those with only two inflections (F = 4.29, df = 3/117, p = .01). For the latter sentences there seemed to be a sharp initial decrease in the Ss' ability to recall the CVCs, followed by an equally sharp recovery. One possible explanation for that result is that with several uninflected CVCs the Ss should have more difficulty in assigning the right P-marker to the sentence than if they are all inflected, and that might have resulted in some confusion for the Ss. With repeated occurrence of the frames, however, the Ss should improve in their ability to assign the right P-marker and their recall of the CVCs should increase as well.

In general, the results of the experiment seem to support the hypothesis that the facilitating influence of structure on learning verbal sequences is that of reducing interference from prior learning. Imbedding CVC sequences in sentence frames did not result in an overall improvement in the recall of the CVCs, but it did result in a reduction in the interference from previous items in the list. In this connection it is interesting to note that there was a total of 19 instances where S̲s gave a CVC which was incorrect for that item, but had been correct for a previous item. All 19 instances occurred when S̲s were attempting to recall strings.

CHAPTER VI

Conclusions and Implications

The goals of the project were to: 1) develop a set of research techniques to study the behavioral correlates of the utilization of rules in the generation of language behavior; and 2) to use these techniques to study the way Ss use the rules in response generation. As a point of departure, the Miller and Selfridge (1950) experiment was taken as an illustration of the to-be-explained phenomenon.

The research was initiated on the basis of two assumptions. First, it was assumed that one of the important effects of grammar is to define the behavioral units of language. That is, it will define the behavioral units with which Ss will attempt to process language behavior. The second assumption was that one way the concept of behavioral units can be conceived is that of a decision unit. That is, it may be realistic to assume that Ss do not generate response sequences by using the stimulus consequences of one response element as the stimulus for the next element, with that being the only cue that arouses the habit or association that eventuates in the next response element. In place of that assumption, the current research was based on the assumption that there may be certain loci in a response sequence where Ss make the decisions or execute the associations necessary for the production of several future response elements. The segment of elements encompassed by the decisions or execution of associations at these loci would be the units referred to under assumption one above.

Finally, the research capitalized on the now rather well documented finding that S̲s show a marked hesitation to emit partial or erroneous responses during the course of verbal-learning experiments. That is, S̲s show a tendency to emit short responses in an all-or-none manner. Even when S̲s are instructed to guess they tend to either say nothing at all or give the complete response perfectly. In addition, however, there is a tendency for the number of partial responses to go up as the size of the response sequence increases, although even for long sequences there is a marked all-or-none tendency.

Combining the above points, it was initially hypothesized that the grammar should define phrase-size decision units for S̲s during the course of learning grammatical sentences. In addition, if these are decision units there should be a tendency to emit them in an all-or-none manner. It was this which was labeled the unit hypothesis. From the unit hypothesis it was predicted that, during learning, the conditional probability of an error on any element within a phrase, if other elements are correct, should be very low. Conversely, if all the other elements fail to occur, the probability of a correct on the element in question should be very low. Therefore, it was predicted that the conditional probability that an S̲ would make an error on any word within a sentence, given the preceding word was correct, should be very low if both words fall within the same phrase, but the probability should be larger if the two items fall within different phrases. These probabilities represent the probability that an S̲ will make a transitional error (TEP), and it was hoped that the way S̲s use rules to generate language behavior could be studied by examining the variables which influence the pattern of TEPs witnin a sentence.

The first three experiments were designed to test the unit hypothesis by having Ss learn sentences as paired-associate responses to digit stimuli. The TEPs were then computed and, as expected, the interphrase TEPs were significantly greater than the intraphrase TEPs. In addition, however, the intraphrase TEPs were unequal, but their relative magnitudes were predictable from the intraphrase structure of the sentences. It was concluded that the TEP pattern for a sentence reflected not only the phrase-level units of a sentence, but its entire phrase structure as well. The similarity between the pattern of TEPs and the sentence structure was measured by assigning to each word-to-word transition within a sentence the number of levels in the immediate-constituent analysis of the sentence that the transition also represented a constituent transition. The transitions were then rank ordered for both TEP and level and a correlation computed (rho). Rho was used as a measure of similarity between the obtained and predicted patterns, and tne correlation was substantial. The hypothesis that Ss use phrase-size units when learning language sequences received additional support from the finding that phrases are learned more rapidly than nonphrases, even when both sequences are equally probable as sentence segments in the language.

In an effort to account for the relationship between the TEP patterns and the structure, a set of five hypotheses were outlined and examined empirically Of the five hypotheses, the one most consistent with the data was one which assumed that Ss used the phrase-structure rules as a system of decoding rules, the utilization of which being characterized by the tree diagram. Most simply, the hypothesis assumes that Ss decode encoding units, which conform to constructions, into their constituents, storing the second constituent in short-term memory while the first is further decoded, etc.

Experiments 9 and 10 were an attempt to look at two of the more critical implications of the decoding operation hypothesis. Specifically, experiment 9 examined the nature of the associative system that the hypothesis assumes Ss use when they are generating a sentence. There are certain points within a sentence where the S should be able to use a pre-existing association between words, and other points where such associations should not be usable. The results were in accord with the hypothesis.

While the hypothesis assumes that Ss proceed through a system of decoding operations when generating a sentence, there is a great deal of variation in the number of operations at each transition. Experiment 10 was designed to determine whether there were time factors correlated with the variation in number of operations. The results indicate that the latency between the stimulus and the first word of a sentence is directly proportional to the number of operations that the hypothesis assumes that Ss make before uttering the first word of the sentence.

The results of these two experiments offer some inductive support for the decoding hypothesis in that they tend to confirm the expected nature of the associations that Ss use when generating a sentence, and the point at which the associations occur. The next experiment was designed to look more closely at the loci and inclusiveness of the decoding decisions. It was predicated on the assumption that if a response unit is made either longer or more difficult, the behavioral effect will not occur at the point at which the new or difficult item is introduced, but at the point at which the decision is made for that unit. The results indicate that as the size of a unit increases, the increased difficulty is not apparent within the unit, but on the transition to the unit.

The implications of the hypothesis were then extended to a developmental dimension by examining the developmental functions obtained in other experiments which were concerned with the grammatical behavior of children. By combining these results and the decoding-operation hypothesis, an expected function was derived for the way the pattern of TEPs for children should gradually approximate the pattern of adults. The results of the experiments were very close to the expected function.

The final experiment in the series was concerned with determining the way Ss use grammar when they are learning sentences. It was predicted that one of the influences of grammar should be that of reducing interference from prior learning. The experiment used structured vs. unstructured sequences of nonsense syllables in a short-term memory task. The results indicate that while there was no overall difference in the Ss ability to recall the two kinds of sequence, there was a build-up of interference during the epxeriment for the unstructured sequences, but not for the structured sequences. The results were taken as support for the differential interference hypothesis.

One of the major implications of the series of experiments is that *S*s use structure-defined units as decision units when they are attempting to recall sentences. A question which has not been touched is whether the same units are used when *S*s are perceiving language material. Miller (1962, b) suggested that *S*s may use structure in perception by using phrase-size decision or interpretation units. He suggested that *S*s may put phrases, as they occur, in some sort of temporary memory storage. When a grammatical marker occurs which signals the end of a phrase, the *S*s remove what has been stored and interpret it while the next phrase is being stored in the short-term memory.

The nature of the perceptual units that *S*s use was explored in an experiment by Fodor and Bever (1965). They presented *S*s with sentences which had either one or two major constituent breaks within them. There were recorded on one track of a stereo tape recorder. Clicks were recorded on the other track such that they occurred either within the first few syllables on either side of the constituent break or in the break itself. After hearing a sentence the *S*s wrote the sentence and indicated, with a check mark, where they thought the click occurred. They reported that in those instances where the click did not occur at the break there were significantly more errors in judgment than when the click did occur at the break. In addition, they report that that when the click did not occur in the break and the *S*s did make an error, there was a significant tendency for the *S*s to indicate that the click occurred closer to the break than it actually occurred. They interpreted their results as suggesting the psychological reality of the phrase as a perceptual unit.

An attempt was made to replicate that experiment in the present series of experiments. Unfortunately, their results were only supported in part. As in the Fodor and Bever (1965) experiment, the probability of accurately locating the click was significantly greater if the click occurred at the break, than if it did not. The percentages obtained in the replication were almost exactly the same as obtained in the original experiment. However, there was not a significant tendency for the errors in locating the nonbreak clicks to be closer to the break than the click actually occurred. Where Fodor and Bever found between 70% and 80% of the errors to be in the direction of the break, only 51% were in that direction in the replication (with chance being 50%). In an earlier study by Ladefoged and Broadbent (1960) using the click technique, it was found that there was a significant tendency for S̲s to indicate having heard the click prior to its actual occurrence. Viewing that result in terms of the right-to-left ordering of a sentence, their errors tended to move to the left. Taking that result as a lead, the results of the replication were rescored to determine the percentage of errors that move to the left of where the click actually occurred, as opposed to the percentage that move to the right. The results indicate that there was a significant tendency for the errors to move to the left (78%, $p < .01$).

While the lack of agreement between the results of the Fodor and Bever experiment and the replication indicate that the results must be interpreted with caution, it does seem possible to conclude that the results suggest the psychological reality of the phrase as a perceptual unit. That conclusion is supported by the fact that perception seems to be more accurate at the phrase boundary where Miller (1962, b) assumes interpretation occurs.

Viewing the research in general, it does appear that it did meet the goals set for it. The paired-associate learning task, coupled with the TEP scoring, resulted in a rather powerful technique for examining the use of grammatical rules. As a technique it appears to be both sensitive and reliable, and it has been sufficiently flexible to be used to examine a number of different kinds of questions.

Regarding the second goal of the research, the results seem to offer rather striking support for the decoding-operation hypothesis or model as a description of the process of sentence generation.

References

Aborn, M. and Rubenstein, H. Information theory and immediate recall. *J. exp. Psychol.*, 1952, 44, 260-266.

Archer, E. J. Re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.*, 1960, 74, No. 10, (Whole Number, 497).

Chomsky, N. *Syntactic structures*. The Hague: Mouton, 1957.

Epstein, W. The influence of syntactical structure on learning. *Amer. J. Psychol.*, 1961, 74, 80-85.

Esper, E. A technique for the experimental investigation of associative interference in artificial linguistic material. *Lang. Monogr.*, 1925, No. 1.

Fodor, J. and Bever, T. The psychological reality of linguistic segments. *J. verb. Learn. verb. Behav.*, 1965, 4, 414-420.

Frankart, J. On the relationship between grammaticalness and the utilization of short-range associations. Unpublished doctoral dissertation, The Ohio State Univer., 1964.

Glanzer, M. Grammatical category: A rote learning and word association analysis. *J. verb. Learn. verb. Behav.*, 1962, 1, 31-41.

Gough, P. Grammatical transformations and speed of understanding. *J. verb. Learn. verb. Behav.*, 1965, 4, 107-111.

Hockett, C. A manual of phonology. *Indiana Univer. Publications in Anthro. and Ling.*, (Memoir 11), 1955.

Horton, D. and Kjeldergaard, P. An experimental analysis of associative factors in mediated generalization. *Psychol Monogr.*, 1961, 75, No. 11, (Whole No. 515).

Jenkins, J. and Palermo, D. Mediation processes in the acquisition of linguistic structure. In U. Bellugi and R. Brown (Eds.) The acquisition of language. *Monogr. Soc. Res. Child Develop.*, 1964, 29, 141-149.

Johnson, N. The cue value of sentence frames for the acquisition of speech categories. Unpublished doctoral dissertation, Univer. of Minnesota, 1961.

Ladefoged, P. and Broadbent, D. Perception of sequence in auditory events. *Quart. J. exp. Psychol.*, 1960, 12, 162-170.

Lashley, K. The problem of serial order in behavior. In L. Jeffress (Ed.) *Cerebral mechanism in behavior*. New York: Wiley, 1951, pp. 112-136.

Lees. R. Models for a language user's knowledge of grammatical form. J. Comm., 1964.

Keppel, G. and Underwood, B. Proactive inhibition in short-term retention of single items. J. verb. Learn. verb. Behav., 1962, 1, 153-161.

Maclay, H. and Osgood, C. Hesitation phenomena in spontaneous English speech. Word, 1959, 15, 19-44.

Maclay, H. and Sleator, M. Responses to language: Judgments or grammaticalness. Int. J. Amer. Ling., 1960, 26, 275-282.

Marks, M. and Jack, O. Verbal context and memory for meaningful material. Amer. J. Psychol., 1952, 65, 298-300.

Marks, L. and Miller, G. The role of semantic and syntactic constraints in the memorization of English sentences. J. verb. Learn. verb. Behav., 1964, 3, 1-5.

Mathews, E. Developmental changes in the functional knowledge of phrase-structure rules. Unpublished masters thesis, The Ohio State Univer., 1965.

Miller, G. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychol. Rev., 1956, 63, 81-97.

Miller, G. Decision units in the perception of speech. IRE Trans. Information Theory, 1962, IT-8, 81-83. (b)

Miller, G. Some psychological studies of grammar. Amer. Psychol., 1962, 17, 748-762. (a)

Miller, G., Galanter, E., and Pribram, K. Plans and the structure of behavior. New York: Holt, 1960.

Miller, G. and Selfridge, J. Verbal context and the recall of meaningful material. Amer. J. Psychol., 1950, 63, 176-185.

Osgood, C. A behavioristic analysis of perception and language as cognitive phenomena. In Contemporary approaches to cognition. Cambridge, Mass.: Harvard Univer. Press, 1957, pp. 75-118.

Osgood, C. On understanding and creating sentences. Amer. Psychol., 1963, 18, 735-751.

Postman, L. and Riley, D. A critique of Kohler's theory of association. Psychol. Rev., 1957, 64, 61-72.

Prentice, J. Semantics and syntax in word learning. Unpublished doctoral dissertation, Indiana Univer., 1964.

Razran, G. Stimulus generalization of conditioned responses. Psychol. Bull. 1949, 46, 337-365.

Rosenberg, S. Recall of nonsense sentences as a function of grammatical cues. Paper read at Psychonomic Society, Niagara Falls, Canada, 1964.

Rubenstein, H. and Aborn, M. Immediate recall as a function of degree of organization and length of study period. J. exp. Psychol., 1954, 48, 146-152.

Russell, W. and Jenkins, J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff word association test. Tech. Rep. No. 11, Contract M8-ONR-66216, Office of Naval Research and the Univer. of Minnesota, 1954.

Smith, K. Recall of paired verbal units under various conditions of organization. Unpublished doctoral dissertation, Univer. of Minnesota, 1963.

Underwood, B. and Schulz, R. Meaningfulness and verbal learning. Chicago: Lippincott, 1960.

Yngve, V. A model and an hypothesis for language structure. Proc. Amer. Phil. Soc., 1960, 104, 444-466.

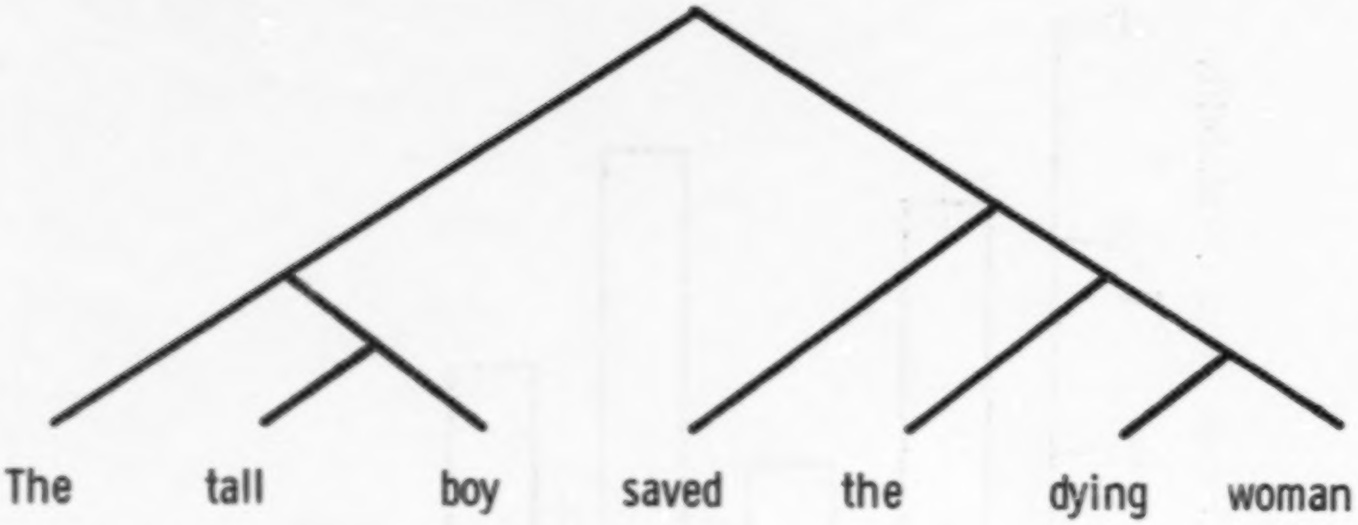

The
tall
boy
saved
the
dying
woman

Figure 1

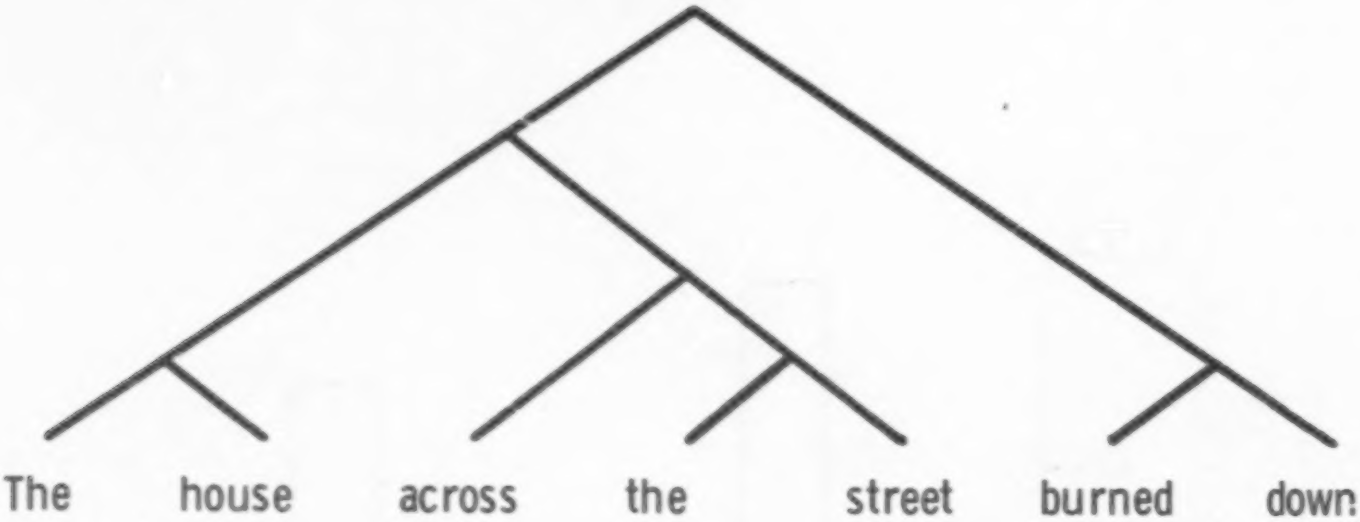

The
house
across
the
street
burned
down

Figure 2

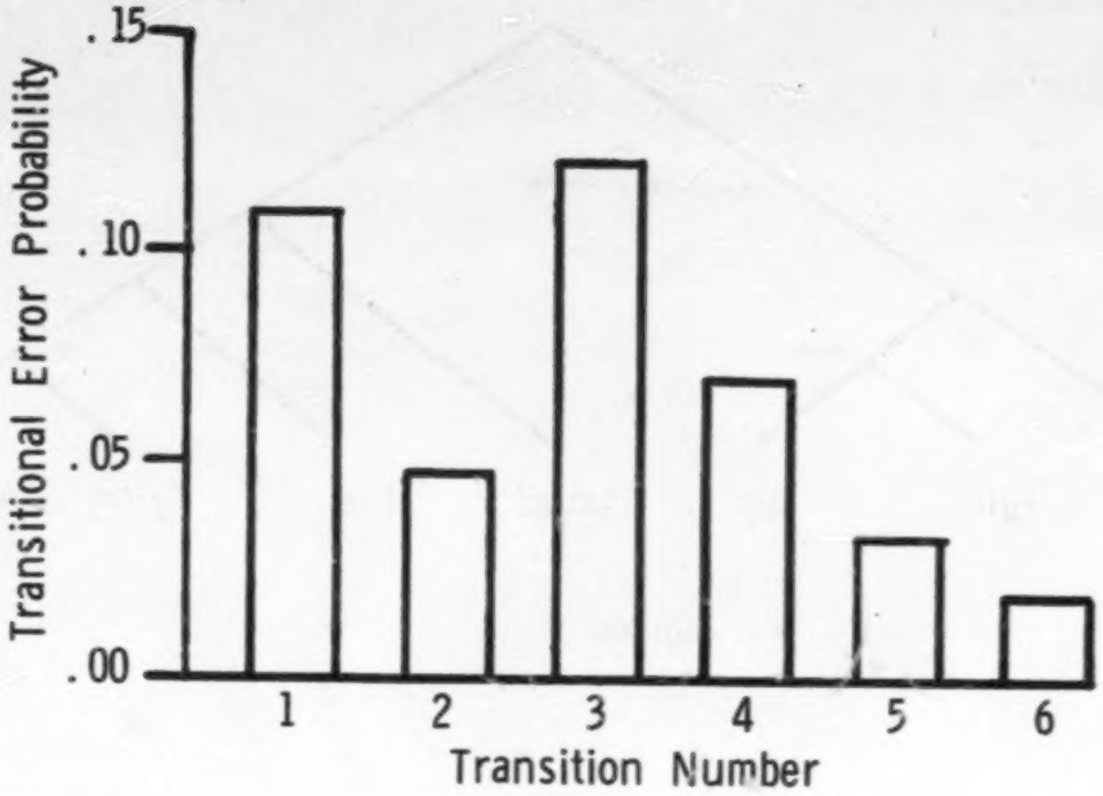


Figure 3

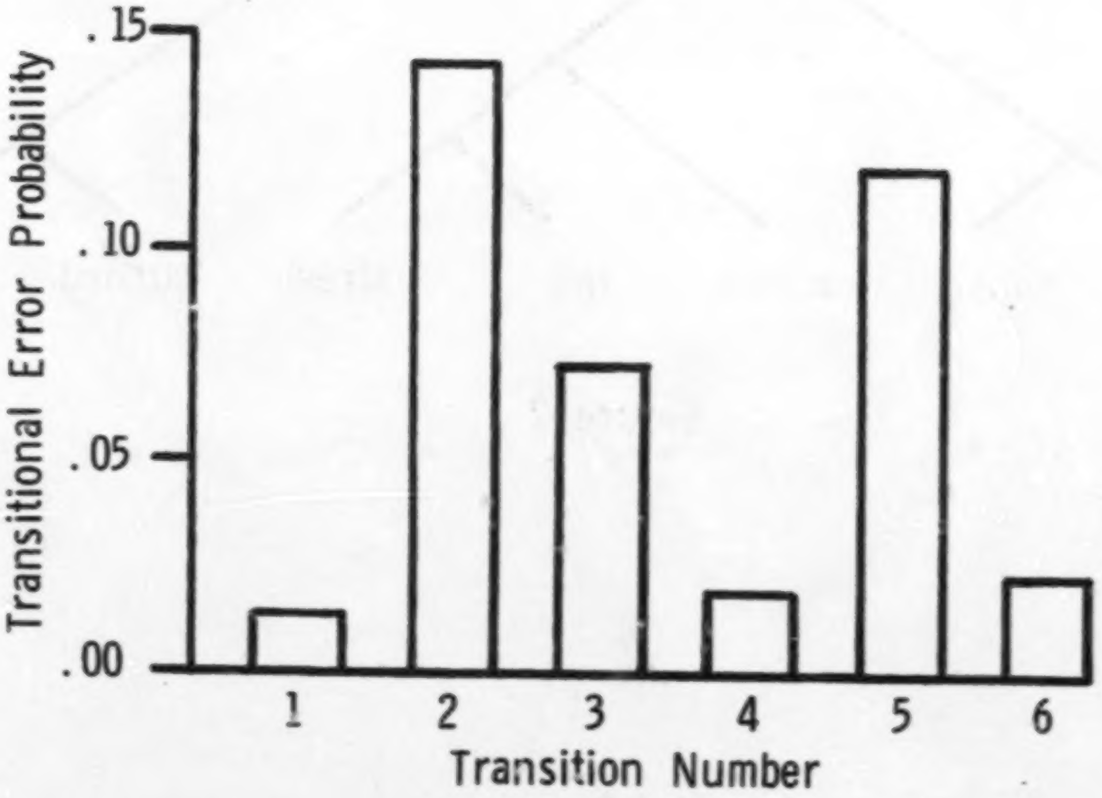


Figure 4

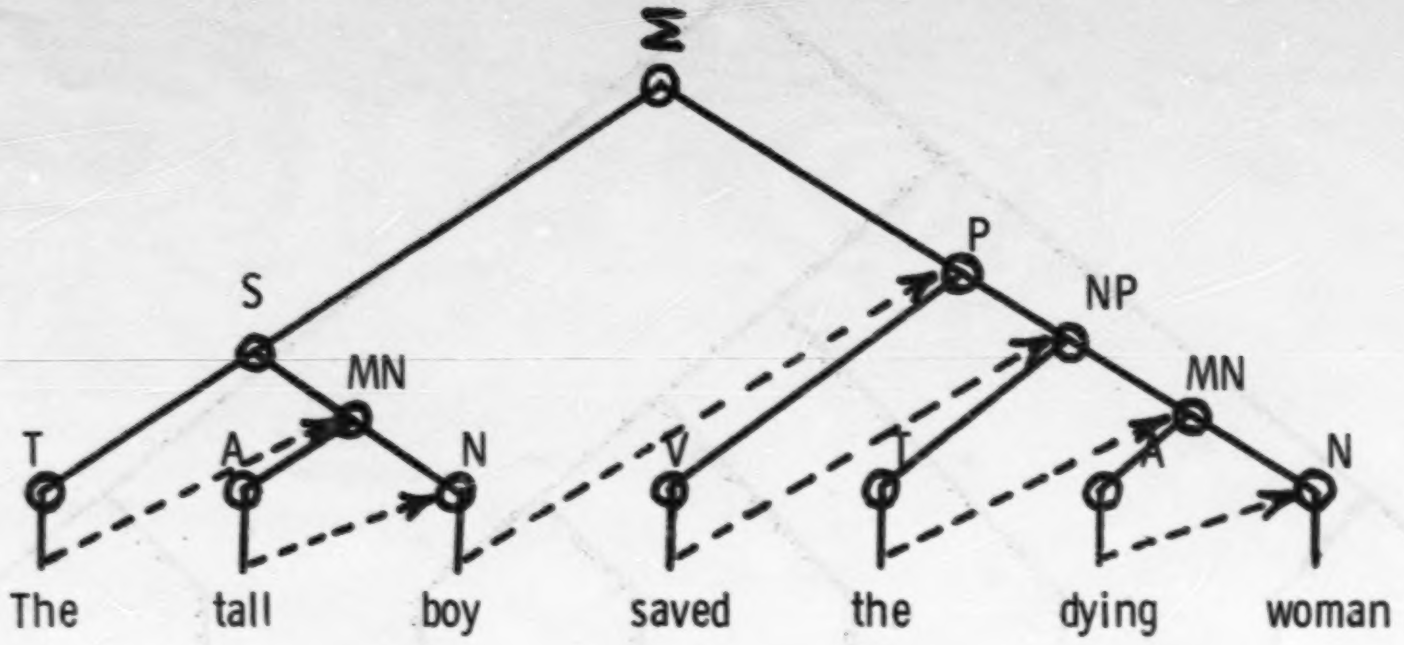

M
P
S
NP
MN
MN
T
A
N
V
T
A
N
The
tall
boy
saved
the
dying
woman

Figure 5

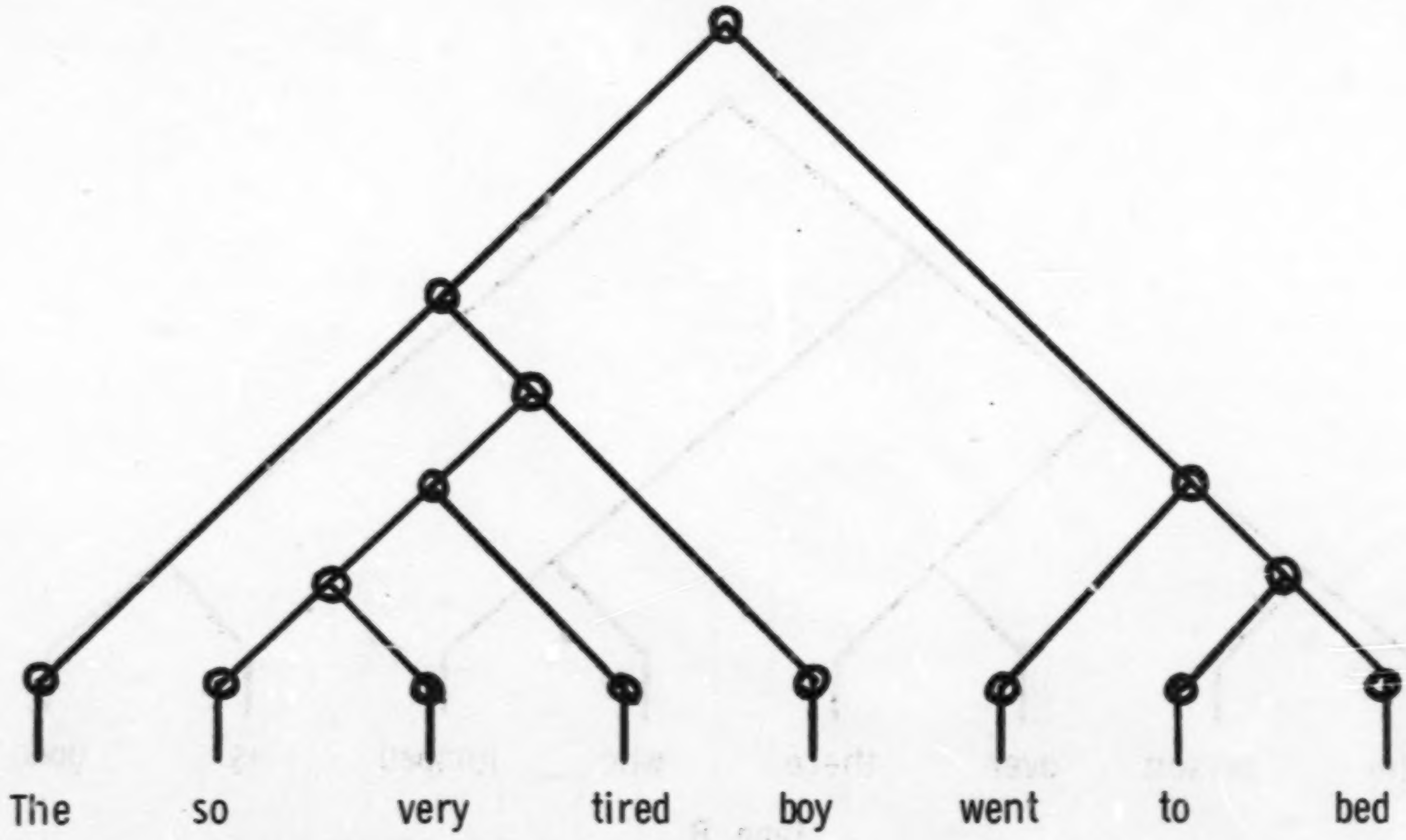
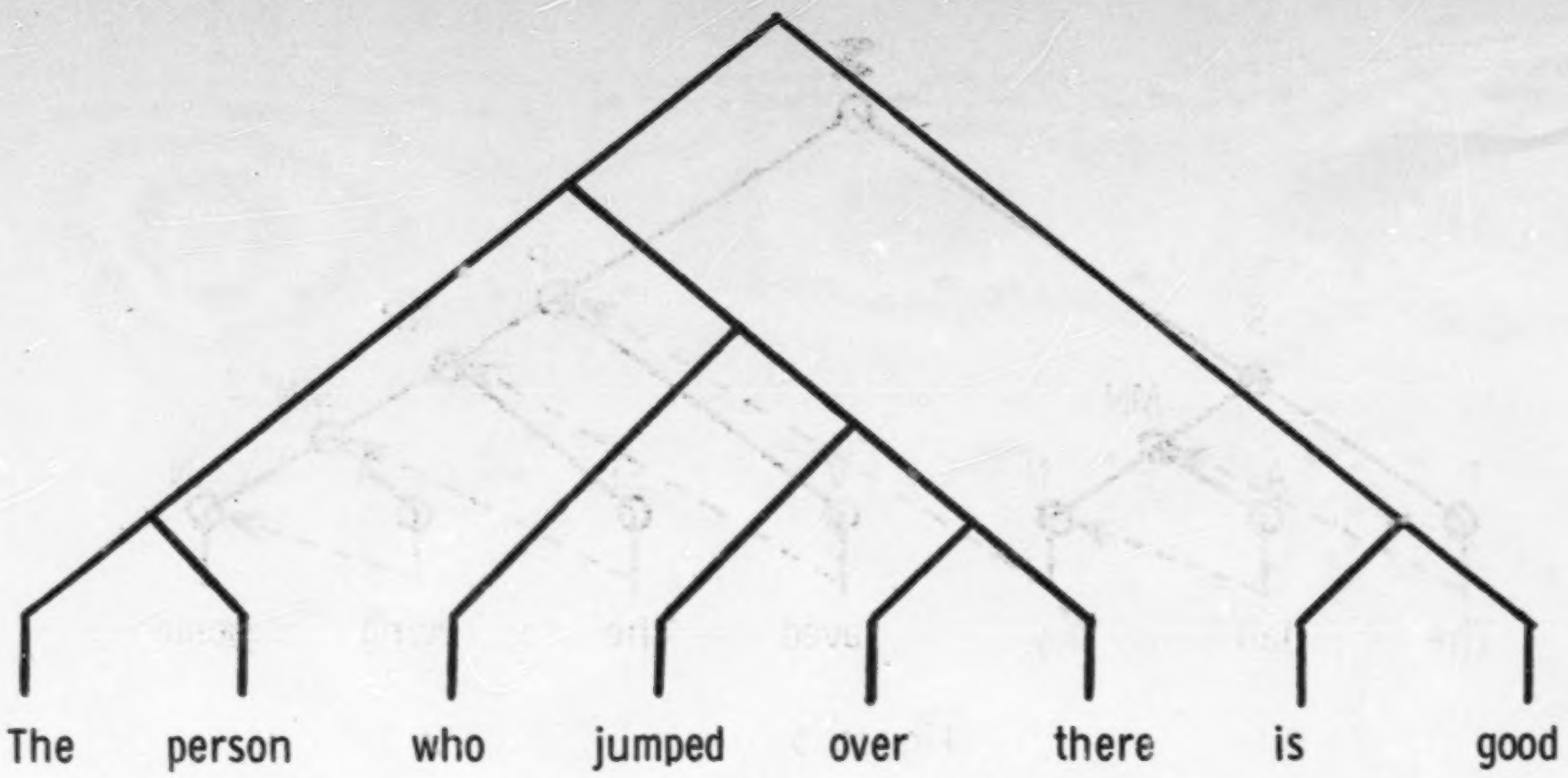
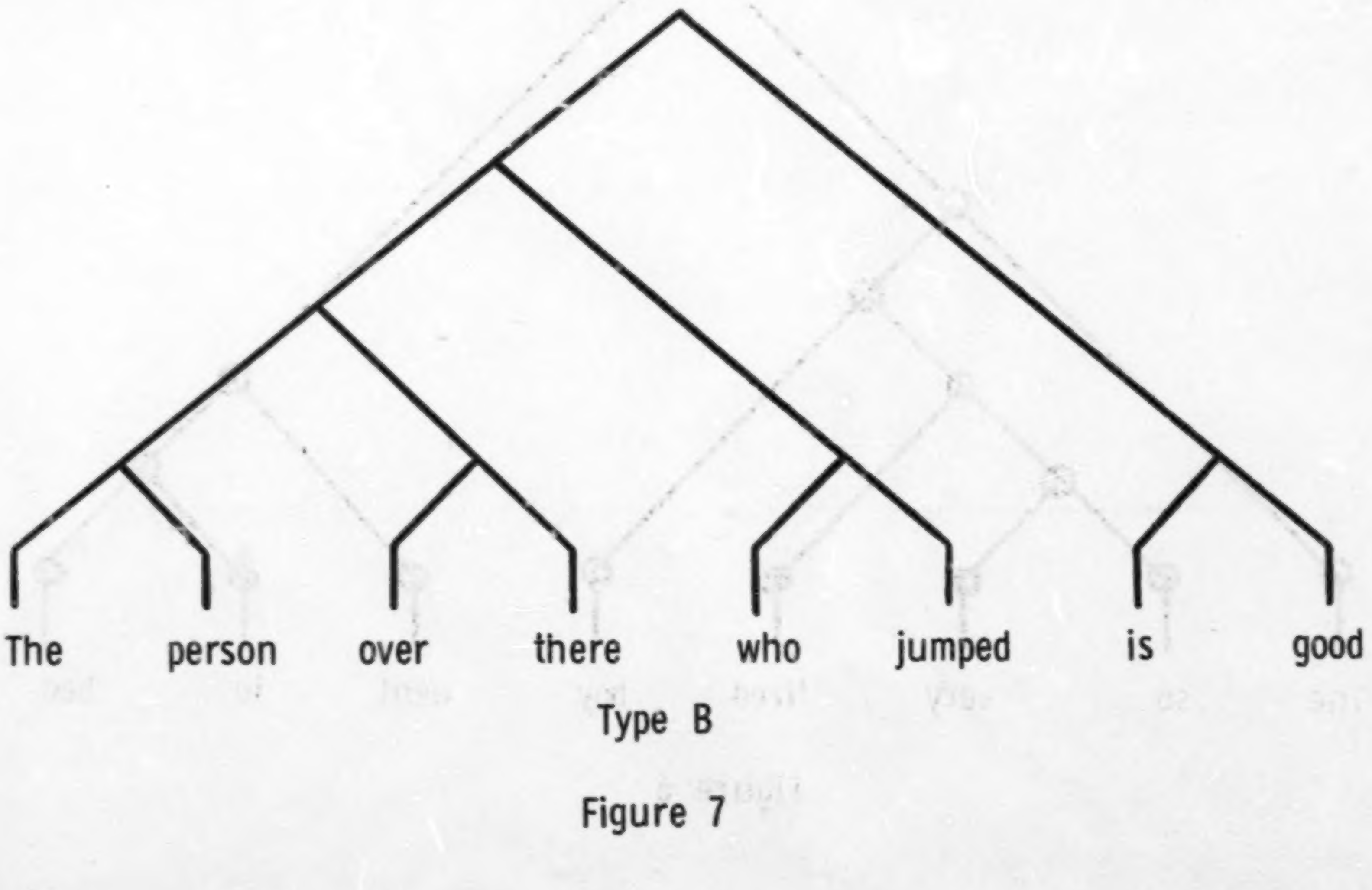

The
so
very
tired
boy
went
to
bed

Figure 6


The
person
who
jumped
over
there
is
good

Type A


The
person
over
there
who
jumped
is
good

Type B

Figure 7

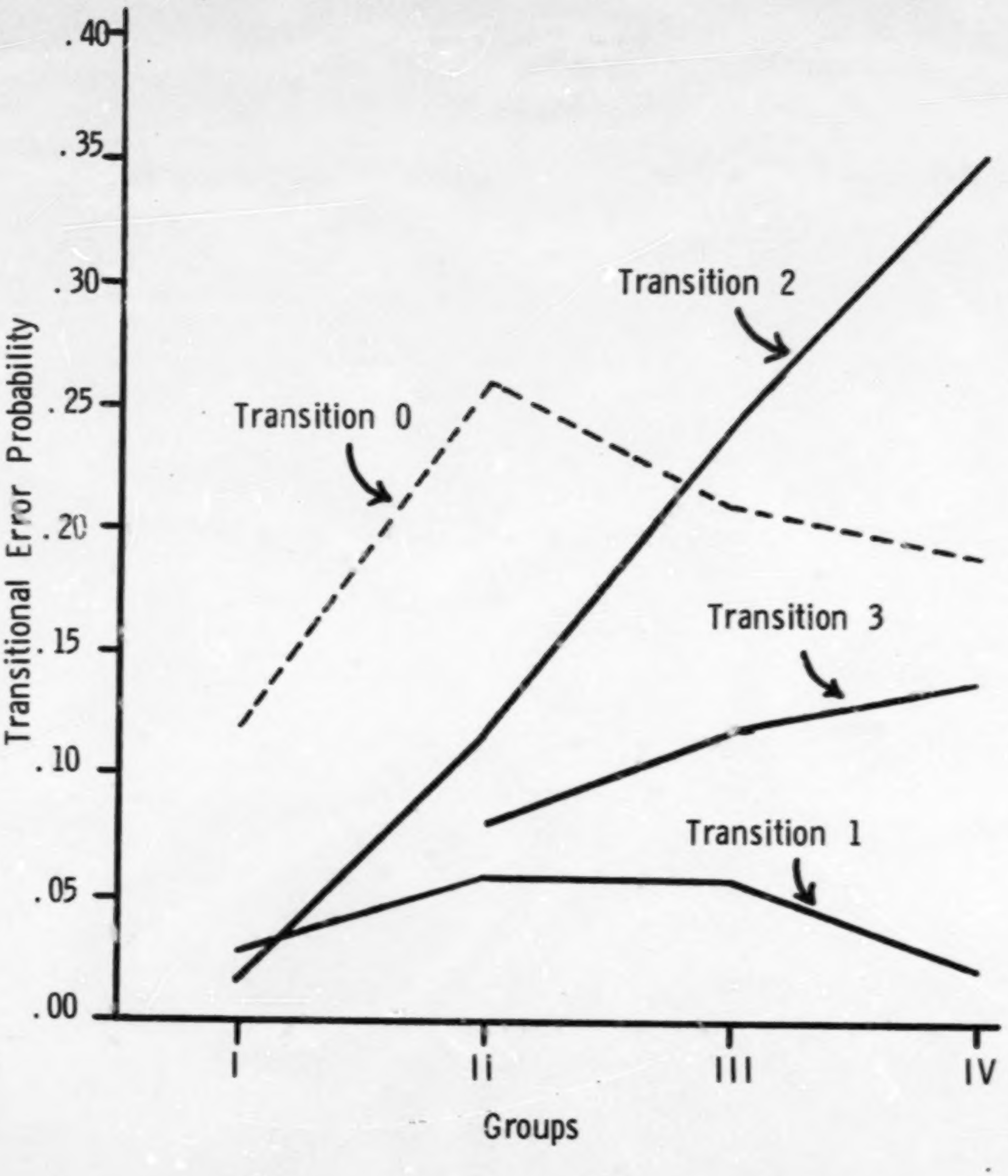

.40
.35
.30
.25
.20
.15
.10
.05
.00
Transitional Error Probability
Transition 2
Transition 0
Transition 3
Transition 1
I
Ii
III
IV
Groups

Figure 8